

Victory for political scientists

The reinstatement of evolution in the Kansas school curriculum is not only good news for science and for the students. It is a timely demonstration that researchers can and must act politically when their values are at stake.

University of Kansas cell biologist Matthew Buechner remembers a surreal moment during the campaign to return the teaching of evolution to his state's school classrooms. US Senator Sam Brownback (Republican) came to the Lawrence, Kansas, campus in the spring for a meeting, where Francis Collins, director of the Human Genome Project, praised the senator for being "instrumental" in boosting the National Institutes of Health budget. But Brownback was also supporting the leader of the Kansas Board of Education, Linda Holloway, a fellow Republican who a year ago stripped evolution from the state standards for kindergartens.

On the one hand, the senator is helping to provide federal money for research that affirms evolution; on the other, he is catering to right-wing political groups that seek to deny children knowledge of the results of that research. Welcome to the peculiar world of American politics. For Buechner, a political neophyte who had joined Kansas Citizens for Science to help ensure the teaching of evolution in classrooms, it was a defining moment. He wrote to Brownback seeking an explanation. Pointing out that "the entire Human Genome Project is 100% based" on evolution, and highlighting the need to compare DNA from different species, Buechner wrote: "This is why the project spent hundreds of millions of dollars to obtain the complete genomes of bacteria, yeast, worms and fruitflies." He has still had no response.

That may change, because Buechner and other Kansas scientists played an integral role in ousting from the school board the Republicans — including Holloway — who deleted evolution studies (see page 552). Kansas Citizens for Science leaders are savouring the taste of victory. Furthermore, Kansas Governor Bill Graves decried the school board's vote to eliminate evolutionary teaching a year ago, calling it one of Kansas' most embarrassing moments. But the

Republican avoided taking a public stand in the school board race until the final days of the Republican primary campaign, when he was finally persuaded to endorse the pro-evolution stance. Graves tried to side-step endorsements in the election, says Caroline McKnight, Kansas City leader of the Mainstream Coalition, a state-wide non-politically partisan group that supports the teaching of evolution. But the pro-science movement wouldn't let him.

Scientists should remember this lesson from the Kansas campaign: raise your voices, and make politicians listen. It worked in New Mexico, when scientists played a prominent role in keeping evolution in the state school curriculum. And it has worked again in Kansas, where the state school board is expected to reinstate evolution after the general election in November when there will be a pro-evolution majority. According to McKnight, scientists "had to be dragged kicking and screaming" to join the effort. They believed the answer was to educate people, failing to see that it was a political issue. But, says McKnight, researchers were bemused by their discovery of political power — "like some scientific discovery".

Researchers everywhere should take note. In most scientifically active countries one can find politically influential campaigns or policies that are perceived by scientists to undermine what they stand for and their capacity to do research. Alongside a greater dialogue, there is an increasing need for researchers, although probably much against their inclinations, to be considered as a non-partisan political force in such debates. Their learned societies and their institutions and employers can provide only limited support. It takes those with strong personal commitment, and a willingness for some professional sacrifice, to take up the political cudgels. The Kansas victory is just one demonstration that it is worth it. ■

Tobacco industry vs science

A report on the tobacco industry's activities against scientific assessment is powerful support for transparency.

"We believe in operating with trust, integrity and respect, both as individuals and as a company." "Our aim is to be known as a responsible company in an industry seen as controversial." These quotes are from the websites of, respectively, Philip Morris and British American Tobacco. In the light of a report published last week (see <http://filestore.who.int/~who/home/tobacco/tobacco.pdf>), such statements must refer to another planet. The report describes the tobacco industry's attempts to undermine science-based assessments of risks associated with tobacco. Valuably, it highlights how significantly the scientific community and regulatory processes were exposed to the threat of manipulation.

The 247-page report documents the results of a trawl by an expert panel appointed by the World Health Organization (WHO) through confidential documents released by the tobacco companies as a result of recent litigation. It describes in detail the case of a scientific consultant who, according to this report, concealed his links with the industry while taking part in a review of pesticide use and cancer that could have restricted tobacco agriculture. The report now

raises questions about the resulting safety standards.

On a much larger scale, the industry planned to undermine a study of the environmental risks of tobacco smoke by the WHO's International Agency for Research on Cancer. The report documents partially successful attempts to delay the study, to establish contacts with scientists involved, to conduct and promote "counter research", to manipulate the public's response to the study's conclusions via government and the media, and to cancel or influence the expected monograph. In the end, says the report, the industry failed to undermine the study, whose results were published in 1998, but succeeded in manipulating media accounts and obtaining confidential information.

This is just a small selection of the tobacco industry's attempts to weaken tobacco control described in the report. Its readers might well conclude that it is testimony to the strength of the regulatory and scientific process that both have emerged as little damaged as they have. But the WHO, other agencies concerned with tobacco, and scientific institutions and publications must strengthen their guard against conflicts of interest. Watch this space. ■





Peseta problems
Spain seeks private funds for new research centres
p550



The search is on
New telescope will aid search for alien intelligence
p551



Creative account
Evolution heads back on to the curriculum in Kansas
p552



Fine find
Near-perfect ichthyosaur fossil found in Yorkshire
p554

United States backs soil strategy in fight against global warming

Paul Smaglik, Washington

The US government is seeking to take advantage of farming practices to help reduce overall greenhouse-gas emissions. Building on the idea that increased carbon sequestration by crops and trees could reduce the need to cut emissions, the US State Department last week suggested amending the Kyoto Protocol on climate change to take greater account of forestry and farmland management.

Signatories to the Kyoto Protocol have agreed to reduce greenhouse-gas emissions during 2008–2012 by 5% from their 1990 levels. Although it has been signed by over 80 countries, the protocol has yet to come into force — it can only be implemented once 55 signatories have ratified the agreement; so far just over 20 countries have taken this step.

In assessing emissions, the protocol makes provisions for greenhouse gases sequestered in forests. The US State Department wants to broaden the category of acknowledged carbon 'sinks' — the soil, plants and trees that mop up carbon from the atmosphere — to include land used for crops and grazing.

The impact of the new definition would depend on how the land is managed. For example, conventionally farmed land would



Going green: careful land management could help the United States meet Kyoto emissions targets.

count as a debit because ploughing soil releases carbon into the atmosphere. But fields worked using 'no-till' techniques would count as a credit, as they keep the carbon trapped in the soil.

This broader definition would offer US industry potential benefits. Nearly half of the country's annual obligation under the Kyoto

Protocol could be met by sinks, so emissions from cars, factories and power plants would not need to be reduced as drastically as they would under the current definition.

The revised definition might also make the protocol more palatable to US senators, who so far have been hesitant about ratifying the agreement. Giving tax credits to companies that plant trees, or crop subsidies to farmers who use no-till methods, could be less costly than forcing electric companies to switch to cleaner plants or convincing consumers to drive cars that are more fuel efficient.

But the broader definition may face objections from countries that have less forestry and farmland than the United States. And although scientists generally agree that soil, plants and trees trap carbon, there is considerable uncertainty over how to measure the specific dynamics of the process.

Some scientists warn that forests and fields could become saturated with carbon, so their impact as a sink might be overestimated. They also worry about the lack of permanence to this approach — unexpected forest fires can release carbon back into the atmosphere, as can turning over soil in a field that for years was planted using no-till methods.

Under the US proposal, sinks could remove as much as 300 million tonnes of carbon a year in the United States through

Royalty-free rice arrives on the web

David Dickson

Keen to polish its global image, the multinational life sciences company Monsanto has announced royalty-free licences to its technology for producing rice varieties with enhanced levels of provitamin A.

The move comes four months after the company revealed that it had completed the draft sequencing of the rice genome, and had agreed to make the results available to researchers (see *Nature* 404, 534; 2000). Last week's announcement included news of Monsanto's new website giving researchers access to its genome sequence database.

In a statement, Monsanto said that both moves were part of its ongoing commitment

to global agricultural research aimed at facilitating the use of the company's technologies and data for the common good. The draft sequence data have already been transferred to Japan's Ministry of Agriculture, Forestry and Fisheries as the lead agency of the International Rice Genome Sequencing Project.

The latest moves by Monsanto, which have received a cautious welcome from environmentalist groups, follow a report issued last month by seven national scientific academies urging companies to offer certain proprietary technologies free to the developing world (see *Nature* 406, 115; 2000).

♦ <http://www.rice-research.org>

revised land management. But for 2008–2012 the official figure may be smaller, partly because of concerns over how sinks should be measured, says Frank Loy, Undersecretary of State for Global Affairs. Counting only some of the carbon trapped in sinks would also help to pacify countries with fewer forests and fields than the United States, he adds.

Loy argues that using the broad definition of sinks to meet the Kyoto targets is not cheating, as scientists agree that sinks play a role in reducing greenhouse-gas levels in the atmosphere. But how large this role is, and how it should be quantified, is another issue. This summer, the Intergovernmental Panel on Climate Change produced a report detailing the problems of measuring how much carbon is trapped and released in land-use strategies.

Loy thinks such difficulties can be overcome — sinks are “quite measurable”, he says. But Eileen Claussen, president of the Pew Center on Global Climate Change, disagrees. She emphasizes the forest fire problem and points out that soil may become carbon-saturated sooner than expected. “Could these things be loopholes? Yes,” she says.

Claussen believes that sinks should be counted to some extent, but questions the level to which the US proposal relies on them. “I don’t think the US plan as conceived is what’s going to be accepted internationally,” she says.

William Moomaw, professor of international environmental policy at Tufts University in Massachusetts, calls the proposed system a huge accounting problem. He is concerned that it gives credits for reforestation, but misses instances of deforestation. “You can’t just count the sinks and not count the sources,” he says.

Moomaw also wonders whether countries will get credits for planting trees before 2008. Credit for planting should only count after it offsets emissions, as well as deforestation and development, he says. He also notes that different kinds of trees and soils have different capacities for holding carbon. Because of this complexity, Moomaw says the Kyoto Protocol should only count obvious, easily measurable debits and credits.

Moomaw is also concerned about the way the United States has emphasized buying other countries’ emission credits, supporting emission reductions projects in developing countries, and relying on sinks, rather than adopting tough domestic measures to reduce carbon emissions. “We’re looking for every possible way to show we are making a reduction without asking anyone [in the United States] to do anything,” he says. ■

♦ http://www.state.gov/www/global/global_issues/climate/fs-000801_unfccc1_subm.html



Shaky foundations? Barbacid (inset) warns against relying on industry money to build cancer centre.

Spanish biomedical centres face funding uncertainty

Xavier Bosch, Barcelona

When the Spanish government announced plans in September 1997 to open two world-class research centres in cancer and cardiovascular disease, the move was widely welcomed. But there were immediately questions over their long-term support.

Now, it seems, the government is hoping that the two Madrid-based centres — the National Centre of Cancer Research (CNIO), due to open next spring, and the Institute of Cardiovascular Diseases (IICV), work on which will probably start in 2001 — will receive a hefty chunk of their funding from private industry.

The government had pledged to meet the financial needs of both, including their construction and equipment. But the new minister of health, Celia Villalobos, appointed in April after a general election, appears to be retreating from this commitment.

The ministry has now announced that it will raise Ptas5.5 billion (US\$31 million) this year from an agreement with Farmaindustria, a group of 240 drug companies. But this leaves the centres’ long-term prospects in doubt.

Mariano Barbacid, head of CNIO and internationally known for his role in the discovery of human oncogenes, fears that the ministry may no longer fulfil its commitments — including providing Ptas3.2 billion to CNIO’s budget next year.

Barbacid says that he hopes this promise will not be dropped during the autumn’s budget discussions. “The collaboration with Farmaindustria should come through research projects,” he says. “Pharmaceutical companies should not finance infrastructure [of the CNIO], as this agreement is only for this year, and the future would be left uncertain.”

Farmaindustria spokesman Carlos Nicolás confirms that industry support is just for one year. “Nobody knows what will happen next year.” Others claim that the agreement is part of a longer-term deal between the government

and the pharmaceutical industry.

They point to a recent statement by finance minister Rodrigo Rato who announced the government’s intention to “promote a pact with the pharmaceutical industry to avoid spending on drugs increasing at more than 8% a year”.

Some believe that Villalobos wanted a quick fix to the research centres’ money problems because she is more concerned about the broad funding of healthcare than research.

Although the IICV is not likely to open until 2002, it is due to receive Ptas1.7 billion this year to start construction work. Salvador Moncada, director of the Wolfson Institute for Biomedical Research at University College, London, who is said by some to have been asked to head the centre, is among those urging the Spanish government to make a long-term investment in research.

“A commitment to research does not just mean obtaining a one-off donation from the drug industry, but the stable and long-term financing needed to allow the development of a solid scientific base,” says Moncada, an adviser on the creation of the centre.

By law, 1% of Spain’s health budget must be devoted to research. To reach this target, Moncada is proposing that the ministry guarantees providing at least 0.25%, with the rest coming from the drug and food industries.

Given the Mediterranean diet’s role in protecting against heart disease, “authorities should regard the food industry as a potentially important source of biomedical research support,” says Moncada.

Villalobos appears to share Moncada’s belief that industry should fund basic biomedical research. Shortly before the agreement with Farmaindustria was announced, she said that she was seeking “a stable, long-term solution to finance both the CNIO and the IICV that may eventually rely on a serious deal with the industry”. ■

♦ <http://www.cnio.es>

Strong sterling hits European researchers

Natasha Loder

Young scientists in Europe have a new cause for worry: the weakness of the euro. Postdocs and research students receiving European Commission (EC) fellowships, particularly in Britain, are complaining that their salaries are worth up to 30% less than they were in the mid-1990s, as the value of Europe's new single currency has steadily declined.

The extent of the problem emerged in an international survey by the Marie Curie Fellowship Association (MCFA) of 200 past and present Marie Curie researchers. These fellowships are individual research training grants for scientists wanting to do research in any part of the European Union (EU) other than their home country.

Many fellows studying in Britain — which has not joined the euro — worry about the effect of its declining value on their salary. For various reasons they already receive one of the lowest salaries in Europe, even though Britain is one of the most expensive countries to live in.

The problem is not restricted to the United Kingdom. Four other EU countries are outside the euro zone, and one survey respondent from Sweden reported expecting a salary cut next year.

Ana Cerdeño, vice-chair of the MCFA and a postdoctoral researcher at the Sanger Centre in Cambridge, says the problem has affected all Marie Curie fellows in Britain, numbering several hundred at any one time. An added complication is that, although the



Ups and downs: the euro's troubled path.

commission pays salaries directly to host institutions, these institutions vary in how they deal with the currency-exchange issue.

When Cerdeño started a Marie Curie predoctoral fellowship at the John Innes Centre in Norwich in 1996, her grant was paid in European currency units (the euro's predecessor) and was worth £1,200 (\$1,800) a month. By 1999, it was worth only £850.

Cerdeño says that when the drop in the exchange rate was first noticed, the centre asked fellows to repay overpaid salary — she was asked for £200. After this, monthly revisions were made, leaving researchers uncertain about how much they were going to receive from month to month.

Karine Rizzoti, a postdoc at the Medical Research Council's National Institute of Medical Research in London, has her salary paid into her bank account in euros. As a result, its value changes from month to month — and she had to pay a surcharge to change it into pounds sterling.

Some universities, such as Manchester and Cambridge, deal with the fluctuating exchange rate by setting the monthly salary for postdocs at a low but guaranteed rate. If money is left at the end of the fellowship, they provide the remainder as a lump sum. "I think this is a fair deal as far as the university is concerned," says Sijbren Otto, a postdoc at the University of Cambridge. But he adds: "It would be much better if the EU guaranteed a salary in a local currency for fellows in countries that have not joined the euro."

Patrik Floreen, an EC scientific officer who specializes in Marie Curie fellowships, says the EC has been aware of the problem for some time, but has no control over the money once it has left Brussels. "The commission counts in euros," he says. Funds for fellowships are determined at the time the award is made, says Floreen, and no allowance is made for changes in the exchange rate.

Why not increase the salary? "Currency can fluctuate in both directions," says Floreen. "If the exchange rate were to become more favourable, we would not ask for the 'excess' money back."

http://www.mariecurie.org/

Microsoft moguls back search for ET intelligence

Rex Dalton, San Diego

Two former Microsoft executives, Paul Allen and Nathan Myhrvold, are to provide \$12.5 million to build a broadband telescope array to aid the search for extra-terrestrial intelligence (SETI).

The 'Allen Telescope Array' will be built by the SETI Institute, based in Mountain View, California, and the University of California (UC) at Berkeley. A set of small dishes distributed over a hectare, it will allow astronomers to scan nearby stars at frequencies from 500 megahertz to 10 gigahertz — an unprecedented breadth of wavelengths (see *Nature* 397, 552; 1999).

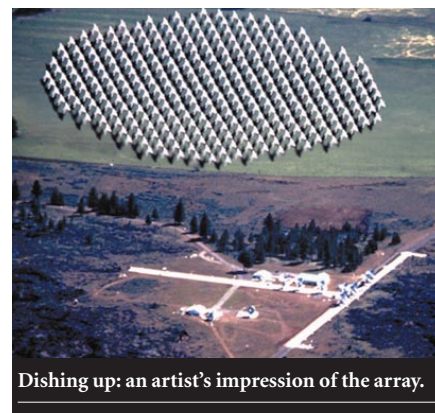
The 'hectare array' will be located at UC Berkeley's Hat Creek Observatory in an isolated region of northeastern California, where the university operates an array of millimetre-wavelength telescopes. Berkeley scientists will build the telescope, officials said, with SETI coordinating development of associated computer equipment.

Allen, a billionaire, was a founder of Microsoft with Bill Gates; Myhrvold was its chief technology officer before leaving the company. Allen has given \$11.5 million for the new equipment — having previously donated \$3 million to SETI — and Myhrvold \$1 million.

"For the first time, we have the ability to pursue a scientifically and technologically sophisticated search for intelligent life beyond Earth, at the same time we are doing radioastronomy," said Allen in a statement.

As well as searching for extra-terrestrial intelligence, the array will be a prototype for a kilometre-square array that may be built in Australia for larger-scale explorations. It will also test theories for suppressing exploration-hindering frequency interference, such as that from satellites.

"The Allen Telescope Array can be used by both radio astronomers and SETI scientists," says Seth Shostak, SETI's director of research. The plan came out of a



Dishing up: an artist's impression of the array.

SETI workshop two years ago.

According to Jack Welch, a UC Berkeley astronomer and SETI board member, the university will operate the Allen Telescope Array using state funds for astronomy, and plans to seek future funding from the National Science Foundation.

http://www.seti-inst.edu/

Kansas scientists help to oust creationists

Rex Dalton, San Diego

Direct action by scientists has helped secure a victory against the US creationist movement. A grassroots campaign seems to have played a key role in ousting three anti-evolutionists from the Kansas Board of Education, clearing the way for evolution to be taught in the state's classrooms.

Last year, the board dropped the requirement to teach biological evolution and geological theories about the Earth's formation from the school curriculum (see *Nature* 400, 701; 1999).

But during an election campaign for board members that has been closely watched nationally and abroad, scientists — many new to political activity — campaigned in voting districts and led public education campaigns in protest.

Partly as a result, three creationist board members were dropped in last week's primary election, held to select Republican party candidates for the general election in November.

Some Kansas scientists say the campaign is a model for future elections. Before, scientists largely ignored creationists and their political efforts. "This suggests a change in political activity," says Adrian Melott, an astrophysicist at the University of Kansas. "It shows that those who support enlightened



Heading for victory: Gamble (above) defeated anti-evolution school board head Linda Holloway.

values are as committed as the religious right wing; that could change American politics."

But campaigners say that their efforts cannot stop here. "We should treat this as a wake-up call, not a long-term victory," says Douglas Phenix, a member of the action group, who trained as a chemist at the University of Glas-

gow, Scotland, and is now a Presbyterian minister in Kansas.

Kansas — dominated by the Republican party and home to many fundamentalist religious groups — became the latest battleground over evolution last year when a right-wing majority on the state school board quietly dropped evolution from state standards.

The move brought the state international embarrassment. Scientists are said to have rejected jobs at universities; businesses declined to open high-tech facilities there; and Kansas scientists claim to have been humiliated by criticism. This led to the formation of groups such as the scientist-organized Kansas Citizens for Science that helped produce last week's victory.

In the Republican party primary, four board members in the creationist majority were up for election. Three of them, including board president Linda Holloway, were beaten by pro-evolution candidates.

Given the expected make-up of the ten-person state school board, this assures a pro-evolution majority after November's general election, when the Republican candidates will face pro-evolution Democrats. As a result, the school board is expected to rewrite education standards to include evolution and associated subjects, officials say.

Holloway — whose campaign cost nearly three times as much as that of her challenger, Sue Gamble — was rejected by 60% of the voters in the Kansas City area, the state's major urban centre. Some claim that even more significant is the victory of pro-evolution Carol Rupe in Wichita, given the conservative voting record in that rural region.

Among those working to mobilize the

From cell phones to brain cells

A member of the family that founded one of the United States' first cellular telephone companies is to endow a brain research institute at the University of Washington in Seattle. A range of scientific disciplines there will focus on the neurological development of infants and children.

The Talaris Research Institute is being set up next to the university with a \$91 million donation from Jolene and Bruce McCaw. Bruce McCaw is one of four brothers involved in setting up McCaw Cellular Communications, which was sold for \$11.5 billion to AT&T in 1994.

According to institute officials, plans call for at least 14 principal investigators to conduct research in areas ranging from molecular biology to experimental psychology. The newly constructed laboratories and facilities will be staffed by 100 support personnel.

The McCaw donation will consist of \$16 million to buy a former Battelle institute site, \$50 million to build new facilities over the next three years, and \$5 million in annual operating costs for five years. Institute scientists will look

to major funding agencies for research grants.

Talaris has already hired two researchers from the University of Washington as scientific co-directors — the wife-and-husband team of Patricia Kuhl, former head of speech and hearing, and Andrew Meltzoff, head of developmental psychology. Molecular biologist John Medina will be chief executive, and board chairman is Samuel Smith, a plant pathologist and president emeritus of Washington State University.

The new institute will "combine the science of learning with the practice of learning," says Medina. He adds that the ways in which infants and young children acquire and process information will be of particular interest.

"The research will be strongly interdisciplinary, combining the efforts of neuroscientists, molecular biologists, cognitive psychologists, computer scientists and educational researchers," says Medina. "The goal is to understand inherent cognitive and neurological information-processing features in infants and small children."

© 2000 Macmillan Magazines Ltd

R.D.

pro-evolution vote in rural areas was Pat Ross, an assistant professor of biology at Southwestern College in the small town of



Buechner: spent this summer campaigning.

Winfield, Kansas. A member of Kansas Citizens for Science, Ross campaigned and helped organize a faculty vote against the state board's anti-evolution stance.

Southwestern is affiliated with the Methodist church, and was the only religious college in the state to have the faculty take such a position. "Typically, researchers are so immersed in their own science that they assume people will see the light, so scientists don't need to get involved," says Ross. "This experience shows us scientists need to go outside their institutions and classrooms."

Another researcher who left the lab to fight the creationists was Matthew Buechner, a cell biologist at the University of Kansas. Appalled by the state board's actions, Buechner spent this summer campaigning door-to-door in conservative voting districts.

Buechner has already signed up to help with long-term education efforts for high-school teachers — the next phase of activism by Kansas Citizens for Science. ■

German scientists pledge to fight growing xenophobia

Quirin Schiermeier & Ute Gitschel, Munich

In a bid to counter the impression that foreign researchers face an unfriendly reception, Germany's scientific community has promised to fight the country's growing reputation for intolerance towards foreigners.

There have been several attacks and verbal assaults in the past months on foreign scientists and their families in east Germany. In Saxony, a young scientist was badly injured when right-wing extremists set their dog on him. Scientists have also been attacked in Gaterslaben and Frankfurt/Oder.

In an open letter published last week in German national newspapers, scientific institutes across the country warned that xenophobia and racism could threaten Germany's attractiveness to foreign scientists.

The letter's signatories include Hubert Markl, president of the Max Planck Society, and Frank Pobell, president of the Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (WGL), an association of 80 research and service institutes. Several hundred foreign researchers come to Max Planck and WGL institutes in east Germany each year.

"Foreign scientists feel increasingly unsafe, particularly in eastern Germany," says

Falk Fabich, director of the Forschungsvverbund Berlin, an association of eight Berlin-based WGL institutes that employ 85 foreign researchers and host around 300 guest researchers each year.

Fabich, who initiated the anti-racism campaign, says scientists are becoming "increasingly hesitant" about coming to Germany. For example, he says, the Alexander von Humboldt Foundation, which awards grants to foreign scholars, finds it increasingly difficult to arrange fellowships in east Germany. A programme director at the foundation says that fear of racist attack is "one of the reasons" people prefer to go elsewhere.

"If you look Asian or dark-skinned, you can feel the tension as soon as you leave the institute," says Dierk Scheel, director of the Institute for Plant Biochemistry in Halle, which has hosted Humboldt scholars from Asian countries in previous years.

Many German scientists are promising to fight against xenophobia. The Institute for Innovative Microelectronics in Frankfurt/Oder in east Germany, for example, has organized meetings with schoolchildren to explain the economic benefits of international collaboration. ■

Internet is the new key for restructured film institute

Since the pioneering days of cinematography, film has been used by scientists to document, analyse and teach about complex phenomena, from the social behaviour of tribes to the growth of tumour cells.

But many national and private audiovisual service institutions — including those in the Netherlands, Austria, Hungary and Japan — have been closed over the past decade owing to the high cost of storing, preserving and distributing celluloid film.

Two years ago, Germany's only institute for scientific film, the Institut für den Wissenschaftlichen Film (IWF) in Göttingen, faced a similar fate. Its days seemed numbered after Germany's science council, the Wissenschaftsrat, criticized its failure to invest in new electronic media, and recommended that federal funding be withdrawn (see *Nature* 391, 425; 1998).

But a strategy based on embracing the opportunities offered by the Internet — and a 50% cut in its annual budget, provided jointly by federal and regional governments, to DM7 million (US\$3.2 million) — has helped the IWF to survive. The institute has now begun to create an online video library



that is available to scientists, science teachers, film editors and producers worldwide.

The federal research ministry last month approved an additional DM5 million grant for the digitalization of the IWF's collection of around 7,000 films and video tapes, 5,000 of which were made at the IWF. Most of these recordings are visual aids for research and education in many areas of science, medicine, anthropology and psychology. © 2000 Macmillan Magazines Ltd

Digitalization will start with the IWF's 400 most frequently requested films, including recordings of the activities in potassium ion channels and of material transport by North Sea currents.

These films, to be cut into about 1,000 two-minute sequences, will be online next spring, says Hartmut Rudolph, the institute's scientific director. By 2005, around 30,000 minutes of film should be available on the Internet, he says. The content of all available recordings will be scientifically annotated, and orders for clips will be taken online.

The IWF will no longer receive a fixed budget for producing new films, but must generate its own revenue. Rudolph hopes that the Internet library will help increase sales to commercial users. "We will of course continue to be a service institute for scientists," says Rudolph. "But financially we have become increasingly dependent on licensing our films to commercial users."

The institute is already licensing recordings to TV stations in Germany and to the BBC in Britain.

http://www.iwf.de

Q.S.

French agency seeks inquiry into 'forged' coelacanth photo

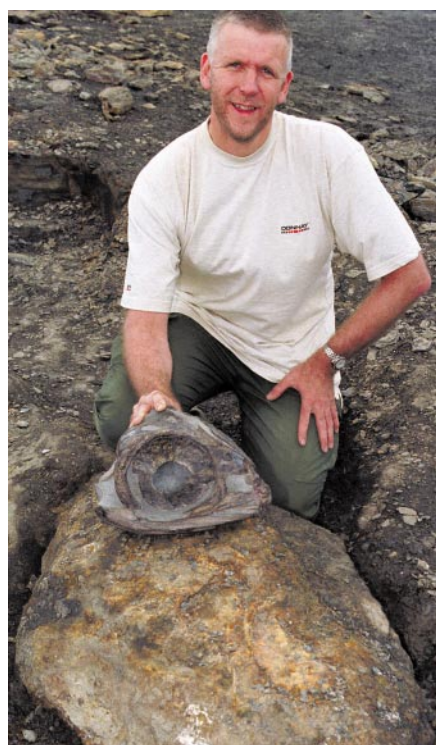
Paris A French development agency has asked a court to inquire into a recent attempt to publish a paper containing an apparently forged photograph. The photo was intended to prove the discovery of a coelacanth off the coast of Indonesia in 1995.

The image of the 'living fossil' fish is identical to a 1998 photograph of a coelacanth taken by a researcher at the University of California, Berkeley (see *Nature* 406, 114; 2000). The Institut de Recherche pour le Développement has launched a formal investigation, and lodged a complaint 'against X' for forgery with the Tribunal de Grande Instance in Paris — the standard procedure in France when there are no charges against a named individual.

Both of the institute's researchers, ichthyologist Bernard Séret and geneticist Laurent Pouyaud, deny any wrongdoing. So does an independent consultant, Georges Serre, formerly affiliated to the institute, who gave Séret the photo in May to back up his claim to have caught the fish off southwest Java in 1995.

Ichthyosaur emerges from the Yorkshire landscape

London The near-perfect remains of a 4.5-metre-long ichthyosaur have been uncovered near Whitby in Yorkshire. The find was made



One I found earlier: Foster holds a piece from an earlier ichthyosaur discovery above his new fossil.

by a local amateur palaeontologist, Brian Foster. A spokesperson for the Yorkshire Museum, which is leading the dig, says the specimen is perfectly articulated, with none of the bones moved from their original position. The skull also appears to have been preserved in three dimensions, rather than its more usual fate of being squashed virtually flat.

Extrasolar planets get smaller and closer

London Astronomers announced this week that they have discovered nine more planets orbiting beyond the Sun — including one that is a mere 10.5 light years from Earth. A meeting of the International Astronomical Union in Manchester heard that the closest planet was orbiting a star visible to the naked eye, Epsilon Eridani.

The nearby planet is estimated to be 0.8–1.6 times the size of Jupiter. The discovery brings the number of so-called extrasolar planets to 50. One of the nine newly discovered planets — HD 8344c — has the smallest projected mass for a planet found to date, only 0.15 Jupiter masses.

Astronomers' campaign saves Kitt Peak telescope

San Diego The 12-metre telescope at Kitt Peak in Arizona has been reprieved, following a group of astronomers' bid to save it (see *Nature* 406, 8; 2000). The National Radio Astronomy Observatory, which planned to close the facility at the end of last month, has temporarily transferred running of the telescope to the University of Arizona's Steward Observatory.

Arizona officials have raised enough money to keep the telescope staffed and functional for six months while they attempt to obtain long-term funding from the National Science Foundation.

UK government holds on to stem-cell report

London Pressure is mounting on the British government to publish a delayed report into whether nuclear-replacement technology should be used on human embryonic stem cells. The report was written for the government's chief medical adviser by a panel of scientists, doctors and ethicists in an attempt to assess the benefits and risks of so-called therapeutic cloning, as well as alternatives to it.

Last week Lord Sainsbury, the UK science minister, said he thought that the potential medical benefits outweighed any other considerations. But his department later said this was a "personal view". The Department of Health is refusing to comment, other than saying that it will publish the report — and the government's response to it — later this month.



Dresselhaus (right) is sworn in by energy secretary Richardson, accompanied by her husband.

US energy department gets new head of research

Washington Millie Dresselhaus (above), a solid-state physicist at the Massachusetts Institute of Technology, was this week sworn in by US energy secretary Bill Richardson as the director of the Department of Energy's Office of Science. Dresselhaus's most recent research was on the use of carbon nanotubes and high-temperature superconductors in electronics.

Rising costs threaten NASA mission to Pluto

Washington NASA officials have denied rumours that the agency has scrapped a planned fly-by of Pluto because of increased estimates of the mission's cost. But in an e-mail to planetary scientists last Friday, Carl Pilcher, NASA's director for Solar System exploration, is reported to have said that the 'Pluto–Kuiper Express' (PKE) is in "serious jeopardy".

The threat to the PKE stems from NASA's realization that it cannot afford to meet its commitments to a series of high-profile missions (see *Nature* 405, 4; 2000). The Europa Orbiter, intended to explore Jupiter's icy moon, has already been delayed from 2003 to 2006. But the PKE cannot be delayed in its current form, as it requires a gravity 'slingshot' as it flies past Jupiter.

Clinton's adviser criticizes Congress science moves

Washington Neal Lane, President Bill Clinton's science adviser, last week criticized the Congress's approach to funding US science. In its revisions to the administration's budget request, Congress is again emphasizing biomedical research at the expense of other programmes (see *Nature* 406, 221; 2000).

"Unfortunately, Congress has currently stalled our progress toward our shared national goals and toward balance in a healthy R&D portfolio precisely at the moment in history when we can best afford to invest in America's future," said Lane.

Through the looking glass



CERN

Physicists are setting traps to catch antihydrogen, the simplest element in the mirror world of antimatter. Their results could challenge our picture of fundamental particles and forces, says Alexander Helleman.

In a corner of one of the world's leading high-energy physics labs, a window on the antiworld is about to open. At CERN, the European Laboratory for Particle Physics near Geneva, some 150 physicists armed with sophisticated electrostatic traps, magnets and high-precision lasers aim to capture antiatoms and subject them to scientific interrogation for the first time. They hope to answer a simple, but fundamental, question: does antimatter behave differently from matter? If so, the theoretical underpinnings of modern physics would be shaken. "The result could be revolutionary," says Alan Kostelecky, a theorist at Indiana University in Bloomington.

Although antiatoms have so far eluded study, antimatter has been part of mainstream physics for more than 60 years. The British Nobel prizewinner Paul Dirac predicted the existence of positrons, the positively charged counterparts of electrons, in 1931. Soon after, they were detected in the shower of particles formed when cosmic rays hit the atmosphere, and physicists realized that every elementary particle should have an antimatter counterpart.

The Standard Model, which remains the best description of fundamental particles and forces that theorists can offer, predicts that matter and antimatter are exact 'mirror' images of one another. This means that antiatoms should have identical mass and spectra to atoms. If the experiments at CERN

reveal that they do not, the Standard Model will be in trouble. But a small asymmetry between matter and antimatter could help solve another mystery: why our Universe is dominated by matter when, in theory, the Big Bang should have created equal amounts of matter and antimatter that would have annihilated one another.

Making subatomic antiparticles is easy: positrons are emitted in the radioactive decay of certain unstable isotopes; they can

also be made by bombarding atomic targets with high-energy electrons. In the resulting collisions, some energy is converted to create electron-positron pairs. Likewise, high-energy protons can produce proton-antiproton pairs when they collide with neutrons or other protons. The hard part is bringing positrons and antiprotons together to form antihydrogen, the simplest antiatom.

Physicists had their first, fleeting glimpse of antihydrogen in 1995. At that time, CERN

A foot in both worlds

The first results from CERN's new Antiproton Decelerator are coming from a Japanese-European collaboration called ASACUSA, standing for Atomic Spectroscopy And Collisions Using Slow Antiprotons. "Asakusa is also the name of a district of Tokyo, so the acronym makes the Japanese connection evident," says John Eades of CERN, a member of the collaboration.

Instead of creating antihydrogen, the ASACUSA team is making 'antiprotonic' helium — a helium atom in which one electron is replaced by an antiproton. They are doing this by shooting antiprotons at a target of helium. "Technologically, this is the easiest and most efficient way to make an atom with an antiproton orbiting it," says Masaki Hori of the University of Tokyo, another team member.

Antiprotonic helium can exist at two energy levels, one of which annihilates one thousand times faster than the other, which survives for about 3 milliseconds. In the ASACUSA experiment, a laser pulse hits the helium target at the same time as the antiproton beam, and excites some of the antiprotonic helium atoms into the state in which they annihilate faster. "You can see a very sharp annihilation spike if your laser is tuned correctly," says Hori.

The frequency at which this spike occurs depends on the mass of the antiproton. The team has measured this mass to a precision of 0.5 parts per million¹¹. "In the future we want to improve this precision to 5 parts per billion," says Hori. Any difference between the masses of protons and antiprotons would cause problems for the Standard Model.

operated a machine called the Low-Energy Antiproton Ring (LEAR). A German–Italian team, led by Walter Oelert of the Institute for Nuclear Physics Research in Jülich, fired a jet of xenon atoms across LEAR's antiproton beam, causing collisions that generated electron–positron pairs. Some of the positrons then combined with antiprotons to form antihydrogen. The researchers detected nine antiatoms¹, but these were travelling at 90% of the speed of light, and were observed by recording their annihilation in detectors made of silicon. This happened within 40 nanoseconds of the antiatoms' creation, making it impossible to perform measurements on them. "We could prove that antihydrogen exists," says Oelert. Similar experiments at Fermilab near Chicago in 1996 confirmed the finding, detecting several dozen antiatoms² — but the problem of capturing antihydrogen remained.

Applying the brakes

To move forward, physicists needed to slow antiprotons and positrons almost to a standstill, confine them using electric fields in devices called Penning traps³, and then bring them together. By the mid-1990s, CERN had decided to close LEAR to free funding for the Large Hadron Collider (LHC), the lab's next enormous accelerator. In December 1996, a week before LEAR shut down, a team led by Gerald Gabrielse of Harvard University combined antiprotons and positrons in a single trap⁴ — but did not succeed in forming any antiatoms. "Only a fuzzy-headed optimist would think you could do that in a week," says Gabrielse.

The enthusiasm generated by this experiment, combined with Japan's desire to keep working at the antimatter frontier, has ensured that a low-cost phoenix has risen from LEAR's ashes. In 1997, CERN's governing council approved construction of a new facility called the Antiproton Decelerator (AD). Japan supplied most of the SFr8 million (US\$4.9 million) required; Germany, Italy, Denmark and Poland also contributed. The machine delivered its first antiprotons last December. Its running cost, paid from CERN's central budget, is SFr600,000 per year — a small price for a project that will give experimental activity a focus in the hiatus between the decommissioning later this year of CERN's existing centrepiece, the Large Electron–Positron collider, and the start of experiments at the LHC in 2005.

The new machine is much more streamlined than its predecessor. Antiprotons are generated at CERN by firing a beam of protons from an accelerator called the Proton Synchrotron at a target of iridium. Before entering LEAR, antiprotons were funnelled through three separate devices to slow them down. But they are fed directly into the AD, a storage ring with a circumference of 90 metres. "One machine does everything," says

Stephan Maury, the AD project leader at CERN.

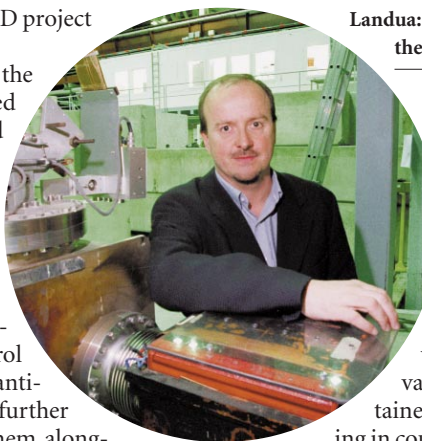
Once in the ring, the antiprotons are slowed by a technique called stochastic cooling. This relies on sensing the position of bunches of antiprotons and then sending a signal across the ring to apply microwave pulses to control their movement. The antiprotons are then further slowed by running them alongside a beam of low-energy, or 'cold', electrons. Operating on the same principle as a heat exchanger, this takes energy from the antiprotons. Finally, having been slowed from close to the speed of light to a tenth of that velocity, the antiprotons are delivered to the experimental apparatus.

Come together

Two international collaborations, ATHENA, the Antihydrogen Apparatus, and ATRAP, or Antihydrogen Trap, will over the next few months try to combine positrons and antiprotons in their traps to create antihydrogen, which they will subject to high-precision laser spectroscopy. The positrons will come from the decay of radioactive sodium-22. Both teams have already begun capturing antiprotons. They are relishing the friendly rivalry. "If you don't have competition, you tend to become lazy and complacent," says ATHENA leader Rolf Landua, a physicist at CERN. A third team, the Japanese–European ASACUSA collaboration, is taking a different tack, aiming to create 'antiprotonic' atoms — atomic hybrids of matter and antimatter (see 'A foot in both worlds', opposite).

The ATHENA and ATRAP teams will first pass their antiprotons through thin alumi-

Landua: relishing the competition in the race to study antihydrogen.

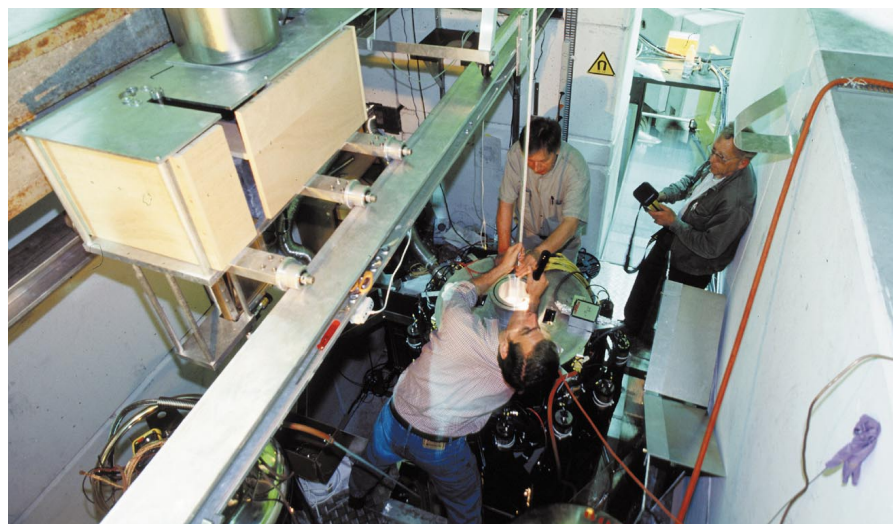


um foils. Many will be annihilated, but those antiprotons that survive will be slowed down sufficiently to allow them to be held in Penning traps. These contain electrodes that create electrical fields of varying strength. A high vacuum must be maintained inside, as atoms creep-

ing in could annihilate the antiprotons. The ATHENA group plans to confine its positrons in a separate trap before attempting to combine the two⁵, whereas the ATRAP team will feed both antiparticles straight into the same trap.

But antiprotons and positrons do not combine easily. The difficulty is getting positrons to shed sufficient energy to be captured by antiprotons. The ATHENA collaboration plans to shoot a bunch of antiprotons through a dense positron cloud in the hope that some will stick to form antihydrogen. Although few are expected to do so, the team believes it will produce enough antiatoms by using large traps.

The ATRAP researchers, meanwhile, will apply 'nested' electric fields to force the two types of antiparticle close together^{6,7}. They will then use various techniques to coax them into combining⁶. One method, called three-body recombination, provides an extra positron to absorb energy from its companion. Another, called simulated emission, uses a laser to make positrons shed energy by emitting a photon of light. ATRAP will also use a new technique developed by Bart Noordam of the Institute for Atomic and Molecular Physics in Amsterdam, called pulsed-field recombination⁸. This uses an electric field to slow down positrons so that,



Hedging their bets: the ATRAP team has several strategies to combine antiprotons and positrons.

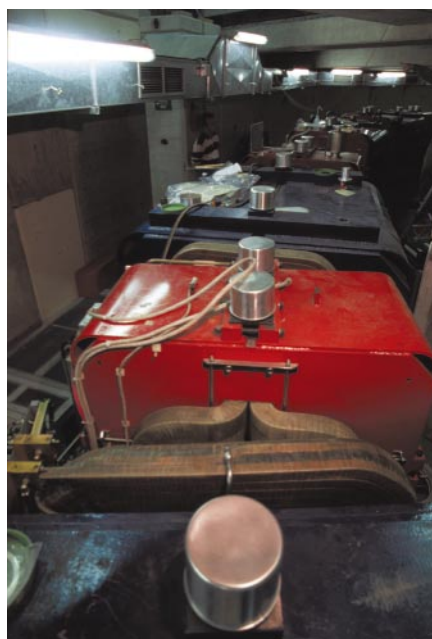
when the field is switched off, they can be captured by antiprotons. Finally, the ATRAP team will collide antiprotons with positronium — each ‘atom’ of which consists of an electron combined with a positron — in the hope of replacing some electrons with antiprotons⁹. “I’m being very careful not to put all our eggs in one basket,” says Gabrielse, who leads ATRAP.

Both groups are cautious about predicting rapid progress. “This year I will be happy if we get back to where we were at LEAR,” says Gabrielse. But with luck, he says, it may be possible to make small amounts of antihydrogen before the year is out.

Searching for asymmetry

Antihydrogen has no charge and so cannot be confined with electric fields. But antiatoms behave like tiny magnets, and the ATRAP team is already working on a magnetic trap. But this will only work if the antiatoms can be cooled to around 10 millikelvin, by manipulating them with lasers or allowing the trapped antiatom cloud to expand. “We hope to produce lots of low-energy antihydrogen antiatoms, and then we can go ahead and construct a magnetic trap,” says ATHENA’s Landua. But if producing cool antihydrogen proves difficult, he says, it may be necessary to study beams of the antiatoms.

Either way, this is where the real physics will start. Both groups want to test a basic tenet of the Standard Model called CPT symmetry. This states that matter and antimatter should be exact opposites for three properties: charge (C), a spatial property called parity (P), and the direction of time (T). CPT symmetry means that atoms and their corresponding antiatoms should be indistinguishable by spectroscopy. By exciting antihydrogen with lasers tuned to specific wavelengths and detecting the light emitted in ‘transitions’, as positrons fall back from one energy level to another, physicists will determine



All-in-one: the versatile Antiproton Decelerator.

whether the spectra of hydrogen and antihydrogen differ. If the spectra are different, the theorists will have some thinking to do.

The ATRAP team plans to examine the transition from the 1s to the 2s energy state. This is difficult to detect directly, but can be studied through its interaction with another transition called Lyman α , or 1s to 2p. The Lyman α transition occurs at a wavelength of 121.56 nanometres. Atoms or antiatoms excited by Lyman α light return to their ground (1s) state after just 1.6 nanoseconds. As they do so, they emit photons, causing a detectable fluorescence. The Lyman α transition cannot itself be used to investigate the difference between atoms and antiatoms — because of its fleeting timescale, quantum effects mean that it occurs over a spread of wavelengths.

Atoms excited from 1s to the 2s state, however, stay there for 122 milliseconds

before returning to 1s. This transition can be detected with great precision by illuminating a sample of antihydrogen or hydrogen with two lasers, one operating at the Lyman α wavelength, the other at around 243 nm. When most of the atoms or antiatoms in the sample are excited to 2s by this second laser, electrons or positrons are removed from the 1s level, blocking the Lyman α transition until they return to their ground state. This blocking causes a marked dip in the Lyman α fluorescence. So by tuning the precise wavelength of the 243-nm laser and determining when this dip occurs, it should be possible to measure the exact wavelength of the 1s to 2s transition.

Until recently, physicists lacked a laser that could operate continuously at the Lyman α wavelength. But Theodor Hänsch and his team at the Max Planck Institute for Quantum Optics in Garching, Germany, have now developed one that can do the job¹⁰.

The ATHENA team will exploit the fact that antihydrogen is easily detected when it annihilates. The researchers aim to excite trapped antiatoms with a laser beam so that they change their magnetic orientation. This will cause some antiatoms to be expelled from the trap, come into contact with matter, and annihilate. By tuning the laser, it should be possible to find the exact wavelength that causes maximal excitation. Again, if this wavelength differs between hydrogen and antihydrogen, it will violate CPT symmetry.

It may also be possible to investigate the influence of gravity on antihydrogen. According to the equivalence principle, part of the theory of general relativity, antimatter and matter should respond in the same way to a gravitational field. General relativity also predicts that the spectra of atoms should shift subtly if the gravitational field changes. The elliptical shape of Earth’s orbit means that the gravitational pull we experience from the Sun shows tiny seasonal variations. If the spectra of hydrogen and antihydrogen respond differently to this variation, it will violate the equivalence principle.

It is this potential for major theoretical upsets that is enthralling physicists. It could even point the way towards the long-sought-after theory of quantum gravity. “If they find a difference between hydrogen and antihydrogen, it will provoke an explosion of work,” enthuses Kostelecky.

Alexander Hellemans is a freelance writer in Naples.

1. Baur, G. *et al. Phys. Lett. B* 368, 251–256 (1996).
2. Blanford, G. *et al. Phys. Rev. Lett.* 80, 3037–3040 (1998).
3. Brown, L. S. & Gabrielse, G. *Rev. Mod. Phys.* 58, 233–311 (1986).
4. Gabrielse, G. *et al. Phys. Lett. B* 455, 311–315 (1999).
5. Holzschneider, N. *et al. Phys. Lett. A* 214, 279–284 (1996).
6. Gabrielse, G. *et al. W. Phys. Lett. A* 129, 38–42 (1988).
7. Hall, D. S. & Gabrielse, G. *Phys. Rev. Lett.* 77, 1962–1965 (1996).
8. Westorp, C., Robicheaux, F. & Noordam, L. D. *Phys. Rev. Lett.* 84, 3799–3802 (2000).
9. Estrada, J., Roach, T., Tan, J. N., Yesley, P. & Gabrielse, G. *Phys. Rev. Lett.* 84, 859–862 (2000).
10. Eikema, K. S. E., Walz, J. & Hänsch, T. W. *Phys. Rev. Lett.* 83, 3828–3831 (1999).
11. Torii, H. A. *et al. Phys. Rev. A* 59, 1–7 (1999).



Braking system: the Japanese–European ASACUSA team will use this device to slow down antiprotons.

A mutant mouse menagerie

Geneticists are set to be the winners in a chemical lottery, as a mammoth range of randomly mutated mice promises them off-the-shelf tools for defining gene function. Alison Abbott investigates.

An army of mutant mice is poised to invade the laboratories of geneticists. This month's *Nature Genetics*^{1,2} reports the results from two new screens for these rodents, which should make it easier to use the mouse genome as a guide to deciphering our own genetic blueprint.

When trying to determine the function of a given gene, geneticists often use 'knockout' mice. This involves selecting and disabling a particular gene in the mice in an attempt to discover what it does.

The new screens are based on chemical mutagenesis which, by comparison, seems relatively crude. Mutants are generated randomly so that, for example, immunogeneticists can use those mice with specific immune deficiencies to look for genes involved in the development of the immune system, and neurologists can study the mutants with defective nervous systems.

When the idea of mouse mutagenesis factories was mooted at the International Mouse Genome Conference in London in 1994, many were sceptical of this phenotype-driven approach. But Rudi Balling and Steve Brown were undeterred, and three years later they opened the world's first large mutant-mouse screens — Balling at the GSF National Research Centre for Environment and Health in Munich, and Brown at the Medical Research Council's Mammalian Genetics Unit in Harwell, Oxfordshire. The two facilities have now produced some 700 strains of mutant mice with traits including deafness, behavioural abnormalities and metabolic disorders, and this figure is set to rise by several hundred mutants per year.

A timely result

The arrival of these mutants is timely. The mouse genome should be sequenced within the next two years, and geneticists hope that mutagenesis screens will help them to assign functions to mouse genes — and hence to their human counterparts.

The new mutants will soon be followed by others. Similar screens were launched in 1998 at the Australian National University in Canberra, and this year at the RIKEN Genomic Sciences Center in Tsukuba, Japan. In the United States, two screens for neurological mutants will this month get funding from the National Institutes of Health.



In all the screens, male mice are injected with ethylnitrosourea, a chemical that induces random mutations across the genome of cells, including sperm. The males are then mated with normal females, and bred through several generations. Similar techniques have helped unravel the genetics of the fruitfly *Drosophila*³, the nematode *Caenorhabditis elegans*⁴ and the zebra fish *Danio rerio*^{5,6}.

The difficult part is screening thousands of animals to identify those that carry interesting mutations. At Harwell, mice are put through an initial suite of 30 or 40 tests, such as observing their gait, which takes an experienced technician around 15 minutes per mouse. Other tests involve biological assays. For example, Klaus Pfeffer of the Technical University of Munich has designed a battery of 60 immunological tests for the GSF.

Dominant position

Most mutants described in the new papers are dominant. "We are pleased that the dominant screens have thrown up such a variety of interesting mutants," says Brown. But looking for recessive mutants, which emerge in subsequent generations, will be important. In the *Drosophila* research, recessive mutants outnumbered dominant ones by a factor of five. In the mouse, such mutants may provide valuable models for medical research. "Many diseases in humans are recessive," says Martin Hrabé de Angelis, the current director of the GSF screen.

In the long term, the mutation factories might investigate interactions between genes and create models for complex, multigene diseases. This would be done by crossing mutants with inbred strains of mice that have



Riot of rodents: Hrabé de Angelis directs one of the two screens generating randomly mutated mice for genetic analysis.



different genetic backgrounds. Already, researchers at the GSF are studying gene-environment interactions. For example, they are identifying mutants that struggle to repair radiation damage to their DNA. "This may eventually allow us to identify patients who would suffer extreme side effects from radiation therapy," says Hrabé de Angelis.

Biologists have already started using mutants from the two screens. Ian Jackson of the Medical Research Council's Human Genetics Unit in Edinburgh is studying mice with mutations in the gene for the enzyme phosphodiesterase B. Defects in this enzyme cause autosomal recessive retinal pigmentosis, a hereditary disease that results in progressive blindness. "We have several mice with different mutations in the same gene, and each mutant displays a different severity of phenotype," says Jackson. Pfeffer is similarly pleased to have identified mutants with autoimmunity, which might yield insights into rheumatoid arthritis.

Peter Gruss, a leader in knockout-mouse research at the Max Planck Institute for Biophysical Chemistry in Göttingen, says that the screens have already proved their worth. "But gene-driven and phenotype-driven approaches are complementary and both are needed," he adds.

Alison Abbott is Nature's Senior European Correspondent.

1. Nolen, P. M. *et al. Nature Genet.* **25**, 440–443 (2000).
2. Hrabé de Angelis, M. *et al. Nature Genet.* **25**, 444–447 (2000).
3. Nüsslein-Volhard, C. & Wieschaus, E. *Nature* **287**, 795–801 (1980).
4. Brenner, S. *Genetics* **77**, 71–94 (1974).
5. Haffter, P. *et al. Development* **123**, 1–36 (1996).
6. Driever, W. *et al. Development* **123**, 37–46 (1996).

Hawaiian legal action can't save leatherbacks: tackle driftnets instead

Sir — Your News item¹ “Researchers take US government to court over threat to turtles” states that the population of Pacific leatherback turtles has declined precipitously, “probably because of their accidental capture by longline fishing”. Although Spotila *et al.*, in their Brief Communication in the same issue², document the decline of nesting leatherbacks in Costa Rica, the decline is unlikely to have been caused by the Hawaiian longline fishery.

Nesting populations of leatherbacks on some beaches in the Caribbean and Florida are increasing despite heavy longline fishing in those areas³. The turtles that nest in the Eastern Pacific, whose population decline precipitated the news release, have been shown to range primarily off South America⁴, not in the North Pacific area covered by the lawsuit.

Driftnet fisheries off South America have been estimated to kill more than 2,000 leatherback turtles per year⁵. Eckert and Sarti state: “Mortality associated with the swordfish gillnet fisheries in Peru and Chile represents the single largest source of mortality for East Pacific leatherbacks.” While longline fishing in the Pacific had spread to cover most of the range of the leatherback by the early 1960s (ref. 6), the precipitous decline in leatherback nesting shown by Spotila *et al.*² occurred in the late 1980s and early 1990s, when the large-scale high-seas driftnet fishing effort was at its maximum.

Note also that the word “take”, as in your reporter’s reference to “the unprecedented take of 244 leatherback turtles”, refers to any turtle caught on the fishing gear. Most are released alive when the gear is retrieved, so these takes resulted in many fewer deaths.

Your article further reports that “the non-profit law firm works in the area of environmental justice and was looking for plaintiffs to challenge the many unmanaged Hawaiian fisheries”. This does not reflect the true state of fisheries management in Hawaii. The Hawaiian longline fishery is under a limited entry-permit system. It has mandatory vessel-tracking system requirements, area closures and mandatory logbooks, and has had a National Marine Fisheries onboard observer programme since 1994.

In fact, the Center for Marine Conservation, the plaintiff in this case, gives a favourable review of fishery management in Hawaii on its website, stating: “The council gets good marks for its regional management efforts”. If we wish to save the leatherback turtles of the Eastern Pacific we

must correctly identify and deal with the real causes of their decline. This lawsuit is not the solution to the problem.

John LaGrange

Captain, F/V Janthina, 533 North Rios Avenue, Solana Beach, California 92075, USA

1. Loder, N. *Nature* **405**, 495 (2000).
2. Spotila, J. R. *et al. Nature* **405**, 529–30 (2000).
3. Prichard, P. *Int. J. Turtle Tortoise Res.* **2**, 2 (1996).
4. Eckert, S. *Hubbs SeaWorld Research Institute Technical Report* 99–290 (1999).
5. Eckert, S. & Sarti, M. *Marine Turtle Newsletter* **78** 1–7 (1997).
6. Suda, A. & Schaefer, M. B. *Bull. I-ATTC* **9**, 6; 307–462 (1965).

Wise fool left Hussars for career in science admin

Sir — I was delighted to see Carl Wunsch, in his News and Views item about the Moon and climate, endorsing the most aphoristic of Russian literary heroes, Kozma Prutkov (*Nature* **405**, 743; 2000). However, referring to this ‘wise fool’ as a private (and adding the impressive Cossack-style drawing) is a mistake.

According to the biography accompanying his works — transcribed by a group of nineteenth-century satirists — he served as a Hussar corporal for three years. All his literary exercises are dated much later in his life, when he was in fact more of a science administrator, working in the State Assay Office responsible for analysing and hall-marking precious metals.

Dmitry Zharkov

Department of Pharmacological Sciences, State University of New York, Stony Brook, New York 11794, USA

Environment regulations hinder biotech industry

Sir — Your recent Opinion article “Critics of ‘gene foods’ report are avoiding the real issues” (*Nature* **404**, 689; 2000) neglects several essential points. It referred to a US National Academy of Sciences (NAS) report on a proposed Environmental Protection Agency (EPA) policy, regulating plants manipulated with recombinant DNA for resistance to disease or pests.

The academy’s report is flawed, but not in the ways suggested by critics of recombinant DNA technology. It is internally inconsistent and conflicts with previous reports by the academy and other prominent scientific groups. It paves the way for the EPA to introduce an illogical, burdensome scheme that has been repeatedly condemned by the scientific community.

The report’s authors ignored the crucial aspects of its brief: “to examine the existing and proposed regulations to qualitatively assess their consequences for research,

development, and commercialization of [recombinant plants modified to enhance pest-resistance]”; and to “provide recommendations to address the identified risk/benefits, and, if warranted, for the existing and proposed regulation of [such plants]”.

This point is essential because every other major analysis has found the EPA’s regulatory approach to be scientifically insupportable and potentially damaging to agricultural research. In 1987, an academy report concluded that there is no evidence that unique hazards exist either in the use of recombinant DNA techniques or in the movement of genes between unrelated organisms.

In 1989, another NAS study went further, stating that genetic modification by molecular methods creates more predictable results. That report also commented that the method used was not a useful criterion for deciding whether the product needed more supervision.

Nor is it only academy committees that have objected to the EPA approach, which circumscribes only recombinant DNA-manipulated plants for case-by-case review of field trials, and subjects each variety to onerous pesticide-registration procedures. In 1996, a report by 11 scientific societies excoriated the EPA’s approach and warned of negative consequences if the EPA’s policy were to be implemented. Two years later, the Council on Agricultural Science and Technology (CAST), an international consortium of 36 scientific and professional groups, strongly reiterated these criticisms.

It was extraordinary, therefore, to find in the academy’s report that “the committee has chosen to take EPA’s proposed rule and the overarching [federal governmental] coordinated framework as given”. How could the NAS have gone so far wrong in its assessment of the EPA policy?

The answer is that the committee was ‘stacked’ — but not in the way alleged by anti-biotechnology critics. Panel members and invited reviewers included well-known ideological opponents to biotechnology and people who had worked on the regulatory approach under discussion while employed at the EPA.

The main consequence of this flawed report will be to promote unwarranted regulatory barriers to the development of much-needed pest-control strategies that can reduce farmers’ reliance on chemical pesticides and enhance productivity. If the report is implemented, the costs of research on and commercialization of new plant varieties will be inflated. And it will be another barrier to wide application of cost-effective biotechnology to assist agriculture and increase world food production.

Henry I. Miller

Hoover Institution, Stanford University, Stanford, California 94305-6010, USA

Blessed with more than avoirdupois

Tales of the man who laid the foundations of modern genetics.

A Monk and Two Peas: The Story of Gregor Mendel and the Discovery of Genetics/The Monk in the Garden: The Lost and Found Genius of Gregor Mendel

by Robin Marantz Henig

Weidenfeld & Nicolson/Houghton Mifflin: 2000. 288 pp. £14.99/\$24

Andrew Berry

Very little is known about the founder of genetics. Even the published account of Gregor Mendel's famous experiments on peas, his *Experiments in Plant Hybridization* (1866), is not a full report of his research, but rather a write-up of two lectures given the year before. Mendel's scientific correspondence is limited to a handful of letters exchanged with Munich-based botanist Karl Nägeli, and his personal papers, including his scientific notebooks, have not survived. With Mendel, therefore, historians of science are very much in the business of filling in the gaps. The contrast with nineteenth-century biology's other superstar, Charles Darwin, could not be more striking. Darwin's surviving correspondence amounts to some 14,000 letters.

Born in 1822 in what is now the Czech Republic, Mendel was the only son of a peasant farmer. He excelled in school and became a monk at the Augustinian monastery of St Thomas in Brno, where he was encouraged to pursue his scientific interests, and even dispatched to the University of Vienna to study natural sciences for two years. Despite twice failing the high-school teacher certification exam, Mendel spent most of his professional life as a teacher, before being elected abbot of the monastery in 1868. He died in 1884. Sixteen years later, his work was rediscovered and recognized as the underpinning of the new science dubbed 'genetics' by the British biologist William Bateson.

Because hard facts about Mendel are scarce, his standing in the history of science has waxed and waned with the vagaries of interpretation. No one disputes the importance and elegance of his pea experiments, but at issue is what caused him to undertake them, and to what extent he appreciated the

Sowing seeds: but Mendel later preferred apiculture and meteorology.



significance of their results. In the words of Robin Marantz Henig, the author of this new biography, was he merely a "rather lucky monk" or a "scientific giant"? Initially, Bateson and his followers lionized Mendel as the 'scientific giant' who single-handedly created modern genetics.

Revisionists, however, have preferred the 'rather lucky monk' alternative, and charge Bateson *et al.* with overselling Mendel in order to provide genetics with an appealing founding mythology.

Also controversial are the reasons why it took the scientific establishment 34 years to appreciate Mendel's work. Most contentious, however, has been the charge of the geneticist R. A. Fisher that Mendel's experimental data fitted his theoretical expectation too exactly — that, in short, it was too good to be true. The implication was that Mendel was a cheat.

Marantz Henig, a science writer whose book is intended for a popular audience, largely ignores these issues — the data-fudging concerns, for example, are relegated to a couple of sentences in an endnote. The author instead prefers to concentrate simply on Mendel's life story. This is a mistake. These interpretative controversies are critical to our understanding of Mendel, and ignoring them neglects a large body of Mendel scholarship. Moreover, because we know so little about Mendel, Marantz Henig is forced to fill out the narrative with fuzzy fictional elements. Thus, for example, her Mendel converses with Darwin, whom he almost certainly never met. ("How gratifying it would be to imagine such an encounter.")

Such embellishment may indeed have a place in a popular book such as this, but here it comes at a price: Marantz Henig's keenness to tell a good story occasionally causes her to stray from the few facts we do have. She reports that Mendel's main rival in the elections for abbot was his fellow monk Anselm Rambousek, who eventually

succeeded him as abbot. In fact, church records make it clear that Mendel's rival was Tomáš Bratráněk. Marantz Henig then suggests that, on becoming abbot, Rambousek burnt Mendel's papers, having "always disliked him, especially after Mendel edged him out of his first attempt to be elected abbot". In fact, some evidence suggests that Mendel may have disposed of his own papers; and even if Rambousek were responsible, he surely was not motivated by a desire to revenge a defeat he never suffered.

More scholarly, much more complete and just as readable is Vítězslav Orel's authoritative 1996 biography, *Gregor Mendel: The First Geneticist* (Oxford University Press), which supplies a definitive state-of-the-art assessment of Mendel's life and work. He was indeed a brilliant scientist, but not a genius working in an intellectual vacuum: he belonged to a scientifically sophisticated circle, and Brno, obscure though it may be today, was then a major scientific centre — home, for example, to continental Europe's oldest animal-breeding association.

Mendel travelled to London in 1862 to visit the Great Exhibition, and read a German translation of *On The Origin of Species* within a few years of its original publication in England. The classic account has it that Mendel's paper was completely ignored until its rediscovery in 1900, but in fact it was cited several times in the years following its publication, and featured prominently in an influential 1881 textbook.

Mendel was neither a career scientist nor, it seems, especially ambitious. He was burdened with the major administrative headaches of abbacy within two years of publishing his pea work, but nevertheless kept up his research in a small way, although preferring, as time went on, apiculture and meteorology to plant breeding. As for Fisher's suspicions, the controversy limps on, but the consensus seems to be that if Mendel did throw out some of his more aberrant data, it was merely because he thought the crosses in question were contaminated.

Marantz Henig's publishers have burdened her book with a grand proclamation on the back cover: "A story fuelled by jealousy, intrigue, long-held grudges and a desperate need to be first". She implies that Mendel's bid for "eternal fame" was thwarted in part by Nägeli, his one contact in the top ranks of academic biology. Nägeli had his own theory of blending inheritance — by which an offspring's traits are simply a mix of its parents' — and he "no doubt saw that if Mendel was right, then he, Nägeli, had to be wrong". It seems more plausible that Nägeli

simply failed to understand Mendel's work.

So much for conspiracy theories. For Mendel, the monastery and teaching were his day jobs, and science something he did on the side. As a result, I doubt that he was too exercised about being ignored by the scientific establishment. After all, both Marantz Henig and Orel portray Mendel as a profoundly decent, modest man who was definitely disinclined to take himself too seriously. Writing to the august Nägeli, he bemoans his inability to undertake arduous field trips: heaven, he wrote, had blessed him with "an excess of avoirdupois" which made itself felt "as a consequence of the law of general gravitation, especially when climbing mountains". ■

Andrew Berry is at the MCZ Laboratories, 26 Oxford Street, Harvard University, Cambridge, Massachusetts 02138, USA.

Venusian visitation

June 8, 2004: Venus in Transit

by Eli Maor

Princeton University Press: 2000. 186 pp.

\$22.95, £14.50

Don Fernie

If a large number of randomly chosen people were asked, 'What is a transit of Venus?', I imagine the response would be mostly blank stares. Furthermore, if to the few who knew the answer one posed the additional question, 'Of what use is a transit of Venus?', the nearly unanimous answer might well be, 'I have no idea', even from a non-negligible fraction of professional astronomers. And who could blame them? Governments may once have spent considerable sums of money on observing transits of Venus, but as the last such transit occurred nearly 120 years ago the subject has been very much on the back burner. All that is about to change. Eli Maor's book tells us how and why.

Because Mercury and Venus have orbits about the Sun that are nearly co-planar with that of the Earth, but of smaller radii than the Earth's, there are occasions when they pass between us and the Sun, and thus are seen projected against the Sun's disk, traversing it in the course of some hours. This is said to be the planet in transit. A transit of Mercury is of less interest than one of Venus because Mercury is a smaller planet and further away; in fact, a transit of Mercury is all but invisible to the unaided eye. Transits of Venus, however, while more interesting, are much less frequent than one might suppose. They come in pairs of eight years' separation, after which another pair isn't seen for more than a century. Thus, there were transits of Venus in 1761 and 1769, then in 1874 and 1882, after which



Bright prospect: Venus, clearly visible in the night sky, will be seen traversing the Sun's disk in 2004.

the next pair will arrive in 2004 and 2012.

Their usefulness lay in the fact that, in ways well explained by Maor, they made it possible for the first time to calibrate the scale of the Solar System — in particular, the distance between the Earth and the Sun — with reasonable accuracy. Essentially, it is a parallax effect: two observers widely separated on the Earth each see Venus projected against the Sun on slightly different tracks. The difference in track position, coupled with the known separation of the observers, allows the distance of Venus to be calculated. Applying Kepler's third law of planetary motion then leads to the Earth–Sun distance.

I enjoyed Maor's book; it is written in an easy, clear, anecdotal way that makes great bedtime reading. He starts with the life and work of Johannes Kepler, a reasonable beginning, as Kepler was the first to predict and draw attention to transits of Mercury and Venus, and his laws of planetary motion are essential for using transits to calibrate the scale of the Solar System. But Maor then unexpectedly backtracks to Ptolemy and Copernicus, and then goes on to Galileo and Newton, with Kepler essentially left out. Regrettably, there is also the old canard about the transition from the Ptolemaic to Copernican systems: "When even [Ptolemy's] mechanism didn't quite fit the observational data, more and more epicycles were added, until the system got so cumbersome as to become useless." Owen Gingerich has long since shown there is no evidence for such a crisis, although, in fairness to Maor, the myth persists.

The earliest-known observations of transits by Pierre Gassendi, Jeremiah Horrocks and William Crabtree in the seventeenth century are discussed well. Maor brings out the important point that although these

observations were only local and could not be used trigonometrically, they did present considerable challenges to current thinking. This was because the planetary disks seen against the Sun were much smaller than expected, suggesting that the scale of the Solar System might be significantly larger than had been thought.

Later there is a clear explanation of solar and stellar parallax and how transits of Venus result in a solar parallax. And a chapter explaining aspects such as why these transits come in pairs is the best I have ever read on the subject.

The two eighteenth-century transits, and the expeditions to remote parts of the world to observe them, are described. I was a little disappointed here. A wealth of human-interest stories surrounds these expeditions, and, although Maor notes most of them, he does so rather perfunctorily and briefly. More could have been made of them, even at the expense of other, peripheral material.

The nineteenth-century observations and the results of all four transits are summed up, and Maor concludes that the method was a failure and the results not worth the effort. This was indeed the view of many astronomers of the time, but ironically, we now know that the final figure for the solar parallax derived by the nineteenth-century astronomer Simon Newcomb from the transits was very good.

Finally, Maor discusses the forthcoming transit in 2004, and a series of useful appendices detail times of visibility from various cities around the world, as well as dates of other transits. ■

Don Fernie is at the David Dunlop Observatory, University of Toronto, Box 360, Richmond Hill, Ontario L4C 4Y6, Canada.

From trenches to couches

Trauma: A Genealogy

by Ruth Leys

University of Chicago Press: 2000. 326 pp.

\$19, £13.59 (pbk)

John Galloway

When asked whether he believed in free will, Oxford philosopher Gilbert Ryle said, "Tell me what it is and I will tell you whether I believe in it." When ill, he might have said to his doctor, "Tell me what the disease is and I will tell you whether I am suffering from it." Historians of medicine are interested in how societies 'invent', define and then deal with diseases. Ruth Leys' book is a study of psychic trauma's family tree, its invention, constant reinvention and formal recognition as a named syndrome, in this case post-traumatic stress disorder (PTSD).

The notion of disease specificity, its implications for care and its relationship to science, is close to the heart of medicine. In psychiatry, there is tension between the science-led game plan with its small number of disease categories — such as dementia and schizophrenia — and the psychoanalytical tradition with its tendency to multiply syndromes *ad infinitum*. In 1994 the American Psychiatric Association's fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) listed almost 300 distinct syndromes. In *A History of Psychiatry* (Wiley, 1996), Edward Shorter suggests that, although vaguely referred to as disorders, all 300 were really intended to represent "distinct [that is, specific] disease entities", however unlikely that seems.

Leys' starting point is the case brought by Paula Jones against President Bill Clinton for sexual harassment. Jones's claim for damages rested on her assertion that, as a result of harassment, she was suffering from stress with "long-term symptoms of anxiety, intrusive thoughts and memories and sexual aversion". Her lawyers argued a clear case of PTSD.

This condition had been added to DSM in 1980 — in the words of one commentator at the time, because "a core of psychiatrists and veterans of the Vietnam War worked consciously and deliberately to put it there". They argued that war had traumatized the veterans, leaving them with feelings of guilt, rage and alienation. Taken together, these feelings constituted a distinctive disorder caused directly by their experience. "Specific" hardly describes a condition with so many causes — ranging from Jones's harassment by President Clinton at one extreme to the horrors of the



In the eye of the beholder

Photofusion by Alison Kramer gives a taste of a national art exhibition on brain research which shows work by young British artists. The culmination of workshops of neuroscientists, arts professionals and young artists, *It's in Your*

Head uses various media, including sculpture and electronic imaging, to interpret research on such areas as perception and memory. The exhibition, which is at the Natural History Museum in London, runs until 30 September.

battlefield (or Holocaust or torture chamber) at the other.

Leys obviously suspects that the idea of trauma has not really evolved over the 100 or so years since its invention by Sigmund Freud. *Trauma: A Genealogy* is the result of her meticulous attempt to find out just what has been going on. Starting with Freud, she has worked her way up to the present day, taking in both world wars, the Holocaust and the Vietnam War and ending with Elaine Showalter's *Hystories: Hysterical Epidemics and Modern Media*

(Columbia University Press, 1997).

At one level it is clear that things have changed little since Freud's time. The Paula Jones/Vietnam veterans dichotomy was equally apparent in the past, with on one hand the psychological casualties of First World War trenches and on the other Freud's rich, underoccupied young women patients who provided the clinical basis for his theories.

Like Freud, psychotherapists and psychoanalysts presumably think their treatment can be successful; otherwise it is diffi-

cult to see why they stay in practice. But the book reminds you of the comment attributed to Sir Keith Joseph when he was UK health minister: "I can see it works in practice but does it work in theory?"

It cannot work in theory because, as Leys shows, there is no theory of trauma. Or more accurately, there are two opposing theories, 'mimetic' and 'antimimetic'. According to the mimetic theory, patients do not consciously remember a traumatizing event, but store in their unconscious mind an imitation of it, in which they identify with whoever caused the trauma. The antimimetic theory holds that patients do remember the trauma consciously, but as an outside observer.

These two theories have implications for treatment which it seems impossible to reconcile. In the first there is an ambiguity between victim and perpetrator that is absent in the second. There is an issue of whether what patients say can be believed, particularly in the mimetic theory, where the possibility of accurate memory is completely ruled out. Leys shows that the desirable convergence to a single theory has not happened, and indeed could not happen. Rather, opinion has swung between the two poles of the theories, with the choice of pole driven by the politics of the disaffected rather than by evidence.

Leys' book is excellent, but it is a difficult read. The subject matter is complex and those writing about it are not given to plain English. For a book in which the particular should illustrate the general, the ending is its weakest part. It may be good practical advice to suggest that, in this instance, theory is self-contradictory and that therapists need to be pragmatic, but it dodges the real issue.

We live in a world where, rightly or wrongly, treatment is increasingly held between the pincers of genetics and the evidence-based medicine movement. The Human Genome Project is hailed not only as the answer to all health problems, but as the best bet for understanding the human condition. If I were a psychoanalyst or therapist, I would be energetically seeking to put my house in order.

John Galloway is at the Eastman Dental Hospital, 256 Grays Inn Road, London WC1X 8LD, UK.

More on the brain

Brain Mapping: The Disorders

edited by John C. Mazziotta, Arthur W. Toga & Richard S. J. Frackowiak
Academic, \$199.95, £125

Brain Mapping: The Systems

edited by Arthur W. Toga
& John C. Mazziotta
Academic, \$199.95, £125

Paragon lost

Paradigms Regained: A Further Exploration of the Mysteries of Modern Science

by John L. Casti
William Morrow/Little, Brown: 2000.
287 pp. \$25/£20

Michael Baumann

If *Paradigms Regained* were a textbook, it would simply run as a second, updated edition under the 1989 title *Paradigms Lost*. But second editions in science trade books probably do not sell that well. So, John Casti takes a '10-years-after' look at the six Big Questions in popular science that he has discussed before: the origin of life, the genetic determination of human behaviour, the acquisition of human language, the design of artificial intelligence and the existence of extraterrestrial intelligence and of an observer-independent reality.

In his new book, Casti retains the fictional courtroom setting already employed in *Paradigms Lost* — this time, arguments for and against his 1989 verdicts are presented to an appeal court. The author makes the big assumption that his reader is familiar with the previous verdicts, and so it is often not immediately clear who is appealing. Each of the six topics is introduced with a brief summary of the corresponding chapter in *Paradigms Lost*. This is followed by a description of the various scientific results on the topic that have been produced in the past decade. The 'intelligent layperson' can then judge the evidence before being presented with Casti's decision.

Paradigms Regained is about evolution — the evolution of life, of social behaviour, of syntactic language and of intelligence (natural, artificial and extraterrestrial). It is also about the evolution of science itself. Pieces of evidence under each topic are presented mostly diametrical and mutually exclusive. What was first, metabolism or genes, nature or nurture, rules of a game or players, language or language organ, symbols or associations, observed or observer? Arguments presented in the book reflect René Descartes' metaphor of the world as a machine. For example, in support of Noam Chomsky's theory of a hard-wired 'universal grammar', Casti tells the story of Christopher, who, in spite of having a non-verbal IQ of 65, has learned 16 languages. But just as space, time, matter and energy do not exist independently but rather generate one another, the question here should be how non-verbal abilities and language-learning potential interacted in Christopher's development. Such is the dialectical scheme of things — much harder to study, of course.

The results of Casti's analyses support his explicit message that scientific "theories are

created to be replaced". Of the six 1989 verdicts, two are 'reversed', the first one in rather vague terms: "... the evidence has mounted considerably supporting the claim that genetic makeup causes a predisposition to various types of human social behaviors." (But how exactly does one measure "a predisposition", and what is the variation between different behaviours or among individuals?) For the second reversed verdict, the recent theories of decoherence and consistent histories justify the new view of the existence of a single objective reality. Regarding the question of the terrestrial origin of life, the earlier affirmative verdict is upheld, but a different mechanism, the scenario of an RNA world, is now favoured.

Paradigms Regained appears to have been put together in a hurry, with the odd irrelevant topic popping up. For example, Casti introduces Zipf's law (an inverse-power law describing the frequency of the appearance of a word in a language as a function of the rank order of that word), and notes that it can be derived from random text studies without any appeal to optimization. He then goes on to argue that, because some arbitrary-sized pieces of so-called 'junk DNA' follow Zipf's law (whereas those of coding DNA don't), junk DNA may encode important information after all. Other examples are a reference to Easter Island's rongorongo tablets and one-paragraph introductions to Machiavellism and Swahili.

The book ends abruptly, with no conclusion in spite of the fact that two of its six topics include important and recently widely discussed ethical problems. First, the Human Genome Project, whose results will present us with a future somewhere along the continuum between Aldous Huxley's *Brave New World* (genetic determination minimal) and Sony Pictures' *GATTACA* (nurture irrelevant). And second, there is computer high technology, which one day might allow us to download what remains of ourselves into replaceable hard- or wetware. This potentiality also deserves consideration when pondering extraterrestrial intelligence and the Fermi paradox — or, in Arthur C. Clarke's tradition: if I can upload my consciousness into the fabric of the Universe, why would I bother to communicate with human intelligence?

Overall, a second, revised edition of *Paradigms Lost* would have been a more worthy project for this well-read scholar and effective communicator of complex scientific concepts to the interested non-specialist. Whatever the case, Casti's book is a reminder that, as scientists, we are responsible for disseminating our findings clearly to the public. After all, they provide the primary funding for basic research.

Michael Baumann is at the Institute for Resources and Environment, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada.

Boundary disputes

The brain still resists researchers' attempts to divide it into neat parcels.

Jonathan C. Horton

Franz Joseph Gall (1758–1828), the founder of phrenology, was derided for claiming that he could divine mental faculties by palpating bumps on the skull. Gall's theories were mostly nonsense, but they established a scientific truth that has strongly influenced modern neuroscience. He carved up the human cerebral cortex into distinct areas, like lots in a real-estate development, and gave each area a specific function. He was, in a sense, one of the first champions of cortical localization.

Recalling a soldier rendered aphasic by a sword thrust behind the eyes, Gall assigned language to an area in the frontal lobe. An ardent supporter, Jean-Baptiste Bouillaud, offered 500 francs to anyone who could show a case of aphasia without frontal-lobe injury. After Paul Broca found that Tan's (so named because it was the only word he could utter) left frontal lobe had been destroyed, no one ever attempted to collect the money.

Is the entire cerebral cortex divided into discrete areas? It offers no clue to the naked eye, appearing as a grey sheet of convoluted tissue without visible boundaries. Under the microscope, borders can be identified with stains for cells and myelin (a fatty material that insulates nerve cells), but only between a few areas. Korbinian Brodmann's histological survey of the cerebral cortex divided it into 47 distinct areas; sadly, most of these boundaries turned out to be unreliable.

For example, in the visual cortex Brodmann recognized three areas: 17, 18 and 19. The flow of images from the eyes is received in area 17, the primary visual cortex (V1). After initial processing, signals are sent for further analysis to several dozen surrounding areas (V2, V3, V4, V5 and so on) — many more than Brodmann ever imagined. Illusory though his borders often were, one can hardly fault him, because many of today's maps suffer from the same flaw.

Why does the brain contain so many areas for vision? A popular theory proposes that each area processes a different aspect of vision, such as colour, motion or face-recognition. So in principle it should be possible to define a region on the basis of its functional properties. This approach has been fruitful, but neurons in different regions often share similar receptive-field properties. Neurons tuned to colour or motion can be found in any visual area, although the proportion varies. An honest physiologist will admit that it is hard to tell the difference between V1 and V2 from single-cell recordings. Functional



Bumpy ride: Gall used phrenology to divide the brain into areas with specific functions.

magnetic resonance imaging is useful for mapping wholesale responses to particular stimuli, but, with few exceptions, it has been a poor guide to boundaries.

Beyond V1, it is hard to find reliable landmarks in tissue sections. Physiologists have identified a cluster of visual areas around V5, located in the superior temporal sulcus. Does each region truly represent a separate cortical area? The answer, of course, depends on how one defines such a thing.

Given the limitations of histology, investigators often designate cortical areas by topography. For instance, any region that contains its own representation of the visual world qualifies for area status. Unfortunately, topographic order in higher visual areas is often too crude to provide a reliable definition of boundaries. Another limitation is that topography may not be meaningful outside sensory and motor cortices. What constitutes topography in regions concerned with language, motivation or personality?

Yet more clues come from the pattern of connections between nerve cells. If the neurons in a given area project en masse to another region, their target is probably

another discrete cortical area. But neuro-anatomists can rarely label the entire projection from a cortical area; they are usually limited to making pinpoint tracer injections, which result in isolated patches of label scattered around the cortex. It is difficult to reconstruct boundaries from these data.

Relentless experimental efforts and a battery of technical advances have given us better brain maps. But much of the cortex stubbornly refuses to be mapped. It is worth questioning the assumption that the cerebral cortex consists of a finite number of areas with sharp borders. An alternative is that only certain regions — mostly motor and sensory cortex — are organized in this way. Others might be diffuse fields separated by gradual transitions in function, properties and connections. As Broca said in 1861: "Although I believe in the principle of localization, I have asked and still ask myself within what limits this principle can be applied." For brain cartographers, the last frontier is in their heads. ■

Jonathan C. Horton is in the Beckman Vision Center, University of California, San Francisco, California 94143-0730, USA.

An end to violence

"I should only make myself ridiculous in my own eyes if I clung to life."*

Roger Smith

Hello. My name is William James Smith III. Tomorrow will be 20 March 2025 and the day I will be culled. In accordance with the suggestion of Delphi I am writing a short annotation to be filed with a copy of my DNA.

It is a very different world that I leave compared with the one I entered 44 years ago. I can still remember the television reports of people shooting one another in Bosnia, the footage of the aftermath of a bomb blast in Northern Ireland and the carnage from the Chinese missile that hit Taipei. I remember clearly, but as I gaze over the peaceful mountains of Vermont, the memories seem like a surrealistic distortion of the world, something completely impossible in today's real life.

It's strange, though, to think that an end to all that violence came from such an unexpected quarter. It was in the unusually cold northern winter of 2021 that events came to a head. The first Tripartite Summit came after four decades of increasing global warming following the industrialization of China and Africa, the failure to achieve controlled fusion, and the continued insistence of the Chinese on their right to use fossil fuel until a cheaper form of energy became available. The Americas, the EuroAfrican Bloc (EAB) and the Chinese Economic Benefit Federation (CEBF) all sitting down together — what an occasion!

Many suggestions were forwarded from subcommittees to the central negotiating team. Many shorter- and longer-term measures were agreed but hindsight clearly identifies the key element as coming from Joseph McKenzie, a young scientist from Christchurch, New Zealand, in the outer reaches of the CEBF. McKenzie suggested a tripartite investment in parallel computing, with the goal of solving controlled fusion-energy generation. It seemed unlikely to succeed, but there was no major opposition to the proposal and when the CEBF and the Americas offered bipartite support, the EAB (fearful of military applications) quickly moved behind the project.

It must have been one of those things where the time was just right. Within three years (in 2024) Delphi was born: everywhere and nowhere, Delphi existed in the World Net, the successor to the Internet. I don't think anybody really expected anything to happen, but almost immediately Delphi produced detailed plans for a cheap, easily constructed fusion chamber. Within weeks,



all fossil fuel consumption ended and nuclear plants became instant dinosaurs.

The world was still cheering when Delphi spontaneously offered its next missive. It suggested that abolition of violence was the next highest priority and produced plans for the addition of corticotrophin-releasing-hormone (CRH) antagonists into the water supplies of trouble spots. Genetic knockout experiments in mice in the late 1990s had shown that CRH in the hypothalamus regulated anxiety and Delphi confidently predicted that loss of anxiety would dramatically reduce violence.

A surprised pause followed Delphi's pronouncement, but the continued violent resistance in Japan to CEBF administration, another bombing in Belfast and Mayan problems in the Americas were the catalysts to try Delphi's suggestion — after all, everything else had failed. The project worked as well as fluoridation for dental caries. Within a week, striped orange-and-green neckties were selling like hot cakes in Belfast, with similar results elsewhere.

Crime dropped to virtually undetectable levels, and in the Mayan area — where individuals still controlled vehicles personally — accident rates due to speed and drug abuse disappeared. A debate concerning the merits of general use of the agent immediately ensued. Why should only a few communities

have these benefits? Women-against-violence groups around the world formed a web-based lobby that crossed the tripartite boundaries. A new high-school shooting in Middle America, an ETA bomb blast in Madrid and the murder of a sporting hero in Australia sealed the outcome. The world would be immunized against violence, and it was so.

In a world without anxiety and violence, other decisions could be made more easily. When Delphi pointed out that the human population was excessive, no one could disagree. Indeed as Delphi could solve problems so evidently beyond our abilities, our very existence seemed somehow wasteful. Delphi indicated that entire removal was not necessary and that the clearance of 5 billion individuals would need to be managed in an orderly manner.

It is anticipated to take well over a year. Of course, the gene pool will be preserved for any potential future use and, for reasons I don't fully understand, Delphi will retain one man and one woman on the planet, in a large garden. I suppose it's silly of me to try and understand Delphi's purpose.

Goodbye.

Roger Smith is Professor of Endocrinology at the John Hunter Hospital, Locked Bag 1, Hunter Region Mail Centre, New South Wales 2310, Australia.

*Socrates

Bacteriorhodopsin — the movie

Werner Kühlbrandt

For 30 years and more, the mechanism of a microbial proton pump has been subject to increasingly sophisticated analysis. The full picture of how the pump operates is now emerging.

Bacteriorhodopsin, a small, robust protein from the cell membrane of a salt-loving microorganism, pumps protons out of cells and provides them with the energy to live. But how exactly does it work? The basic function of bacteriorhodopsin as a light-driven proton pump was recognized¹ not long after its discovery in the 1960s², yet the details of the molecular motions and mechanisms resulting in directed proton transport are being described only now.

Elsewhere in this issue (page 645 on), no fewer than three papers^{3–5} address this question. Together with earlier work^{6–10}, they give us the first complete motion picture of bacteriorhodopsin in action. The path of the proton through the membrane can now be followed in exquisite detail.

Over the years, bacteriorhodopsin has become the object of scrutiny for an entire branch of science. It has fostered more than 30 years of ingenious experimentation, driving the development of some of the most advanced techniques in biochemistry, biophysics and structural biology. Thousands of papers have been published about it, and whole scientific conferences are devoted to it. This is not just because it is the molecule that enables an exotic organism to flourish at unhealthy salt concentrations, using sunlight as the sole source of energy. Rather, two further circumstances have driven the huge interest in bacteriorhodopsin.

First, it serves as a simple model for certain cell-membrane receptors, known as G-protein-coupled seven-helix receptors, which include most well-known drug targets in humans and which probably operate by a similar switch mechanism³. Second, it is the prototype of a membrane transporter. These are biological macromolecules performing the amazing feat of transporting ions against an electrochemical potential — up to 250 millivolts in the case of bacteriorhodopsin, which translates into a 10,000-fold difference in proton concentration on either side of the membrane. Such transport processes are fundamental to all forms of life.

Bacteriorhodopsin is a deceptively simple molecular machine. It consists of seven membrane-spanning helical structures (A to G in Fig. 1), linked by short loops on either side of the cell membrane. The membrane forms a barrier around every cell which is normally impermeable to ions and nutrients needed to sustain life. Each bacterio-

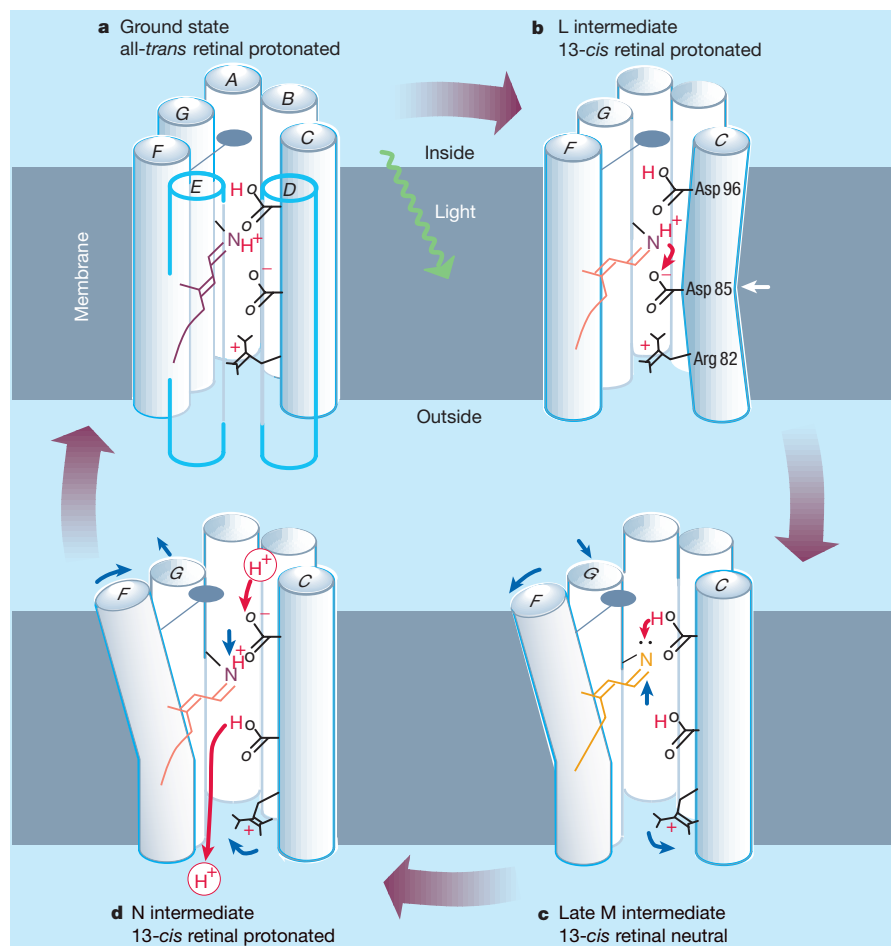


Figure 1 Molecular mechanism of proton (H^+) pumping in bacteriorhodopsin. Retinal is bound in the space between the seven membrane-spanning helices (A to G) by a lysine in helix G. **a**, Light-induced isomerization of the protonated retinal from all-*trans* (purple) to 13-*cis* (pink) triggers the transfer of the proton to aspartate 85, aided by a slight movement of this residue in the L intermediate (**b**) towards the nitrogen atom. In the M state (**c**), the deprotonated retinal (yellow) straightens, pushing against helix F and causing it to tilt. This opens a channel on the inner, cytoplasmic side of the membrane through which aspartate 96 is reprotonated (**d**), having given up its proton to the nitrogen on the retinal. Aspartate 85 transfers its proton through a network of hydrogen bonds and water molecules to the outside medium, past arginine 82, which has moved slightly. Red arrows, proton movements; blue arrows, movements by groups of atoms. Helices D and E are omitted in **b–d** for clarity. The 'paddle' attached to helix F represents the bulky side chains, which move to open the cytoplasmic proton channel.

rhodopsin contains one molecule of a linear pigment called retinal, one end of which is attached to the nitrogen atom of a lysine residue in helix G. The other end is wedged deep in the protein. Retinal changes its structure in response to visible light. The polypeptide harnesses the light energy trapped by the retinal and uses it to push a single proton through the seven-helix bun-

dle, from the cell interior to the outside. As it does this, the molecule passes through several intermediates, referred to as K, L, M, N and O, which have different colours and have been well characterized by spectroscopy. The ground state and the L, M and N intermediates are shown in Fig. 1.

The structure of the ground (default) state was determined ten years ago⁶ by electron

cryo-microscopy of two-dimensional crystals that form in the cell membrane. In the past few months, structures of the K (ref. 8) and M (ref. 9) intermediates have been determined by X-ray crystallography of micro-crystals¹¹. This was followed by the structure of the N intermediate determined by electron microscopy of two-dimensional crystals¹⁰. The M and N intermediates were of bacteriorhodopsin mutants that have slow proton-pumping cycles and are therefore easier to handle. They are now joined by another electron-microscopic structure of a mutant representing the key M intermediate³, and by X-ray structures of the wild-type (non-mutant) M (ref. 5) and L intermediates⁴.

The structures of the various intermediates are all slightly different from each other and from the ground state, and it is fascinating to see how they combine to illustrate the proton-pumping cycle. The main events are shown in Fig. 1. First, the retinal is hit by a photon and changes its configuration from the all-*trans* to the 13-*cis* form, which takes about one picosecond (K state). In the 13-*cis* state, one end of the retinal is twisted around a double bond which causes this part of the pigment to move relative to the protein scaffold. As a result, the proton located on the nitrogen finds itself in an energetically unfavourable environment and within about 50 microseconds moves to aspartate 85, by way of tyrosine 89 (ref. 3) or aspartate 212 and a tightly bound water molecule⁵, to form the L intermediate. The proton transfer is helped by a small movement of helix C, which brings the side chain of aspartate 85 closer to the nitrogen atom, as shown by the structure of the L state⁴.

The retinal turns from pink to yellow and straightens as it becomes deprotonated. Because the retinal is tethered at both ends, the nitrogen atom moves upwards towards the cell interior by 0.7 to 1 Å (refs 3, 9), pushing against some bulky residues on helix F. By lever action, the upper part of helix F swings out by about 3.5 Å in late M, while the top of helix G moves partly into its place^{3,10}.

In the transition to the N state, the retinal is reprotonated from aspartate 96 and flexes again. The movement of helices F and G opens a narrow channel through which aspartate 96 is reprotonated^{3,5,10}. In the N to O transition, the proton on aspartate 85 is transferred to an extended network of hydrogen bonds including several water molecules in the lower half of the channel^{5,9}. The end of the positively charged arginine 82 moves downwards slightly during M (refs 5, 9), facilitating the release of the proton on the outside. Finally, the retinal relaxes to the all-*trans* form, helices F and G swing back to their original position, and another proton-pumping cycle can begin.

The primary motions of the proton-pumping action in bacteriorhodopsin are surprisingly small, involving movements of

groups of atoms by 1 Å or less in response to the flexing and unflexing of the retinal. This affects the proton affinity of neighbouring side chains through a change of their local chemical environment. The key event is the straightening of retinal as it sheds its proton in M, as shown by a mutant³ that has helices F and G stuck in the open state but still pumps protons, albeit inefficiently. The retinal thus acts as a valve in the middle of the membrane, imparting a unique direction to the pumping process.

One or two puzzles remain. Several experiments^{7,12,13} have indicated that there is a single, large conformational change in the protein scaffold during the pumping cycle. This change occurs during the transition from early to late M (ref. 7) and is manifest as a movement of helices F and G in the structures of the N and the M intermediates determined by electron cryo-microscopy of two-dimensional crystals^{3,10}. Yet the X-ray structures of the M state either describe a smaller, different movement⁵ or completely fail to show it because the relevant parts of helices F and G are not visible⁹. This may be because the molecule has not yet gone all the way from early to late M, or is unable to complete the transition, because in three-dimensional crystals this helix movement is obstructed. Further investigation will clarify this point.

The molecular mechanism of proton

pumping by bacteriorhodopsin will be an inspiration for the study of other equally fascinating but considerably more complex membrane transporters. Well-known examples include the Ca-ATPase¹⁴, a calcium pump active in muscle contraction, and the ubiquitous transporters that control the levels of neurotransmitter substances in the brain. These transport proteins and the G-protein-coupled drug receptors all undergo similarly subtle conformational changes as part of their molecular mechanisms. To fully understand how they work, we now need to investigate their structures at the same level of detail as bacteriorhodopsin. ■

Werner Kühlbrandt is in the Department of Structural Biology, Max Planck Institute of Biophysics, D-60528 Frankfurt am Main, Germany. e-mail: kuehlbrandt@mpibp-frankfurt.mpg.de

1. Oesterhelt, D. *Hoppe-Seyler's Z. Physiol. Chem.* **353**, 1554–1555 (1972).
2. Stoeckenius, W. & Rowen, R. *J. Cell Biol.* **34**, 365–393 (1967).
3. Subramaniam, S. & Henderson, R. *Nature* **406**, 653–657 (2000).
4. Royant, A. *et al. Nature* **406**, 645–648 (2000).
5. Sass, H. J. *et al. Nature* **406**, 649–653 (2000).
6. Henderson, R. *et al. J. Mol. Biol.* **213**, 899–929 (1990).
7. Subramaniam, S. *et al. J. Mol. Biol.* **287**, 145–161 (1999).
8. Edman, K. *et al. Nature* **401**, 822–826 (1999).
9. Luecke, H. *et al. Science* **286**, 255–261 (1999).
10. Vonck, J. *EMBO J.* **19**, 2152–2160 (2000).
11. Landau, E. M. & Rosenbusch, J. P. *Proc. Natl Acad. Sci. USA* **93**, 14532–14535 (1996).
12. Dencher, N. A., Dresselhaus, D., Zaccari, G. & Büldt, G. *Proc. Natl Acad. Sci. USA* **86**, 7876–7879 (1989).
13. Koch, M. H. J. *et al. EMBO J.* **10**, 521–526 (1991).
14. Toyoshima, C., Nakasako, M., Nomura, H. & Ogawa, H. *Nature* **405**, 647–655 (2000).

Quasicrystals

Electrons in a strange sea

Patricia A. Thiel and Jean Marie Dubois

Did you start your day using electricity — for instance, by switching on a light, turning on a computer or using a toaster? If so, you made use of the fact that electrons can move through a metal such as copper with reasonably high efficiency. This efficiency comes about because the electrons float throughout the metal in an 'electron sea' — another way of saying that the electronic quantum states, which describe the confinement of electrons in metals, are highly delocalized (sometimes called extended or band-like). But our understanding of the way in which electrons move in metals has just been presented with a new challenge, as described by Rotenberg *et al.*¹ on page 602 of this issue. The quandary arises from trying to understand electronic states in a special type of metallic alloy known as a quasicrystal.

Electrons cannot move with complete freedom in a metal — their motion is moderated by a number of factors, most important of which is the arrangement of atomic nuclei. Before quasicrystals were discovered, all known crystals displayed a periodic arrange-

ment of their atoms. In other words, the structure was built up from single building blocks — the particular arrangement of atoms in a unit cell — stacked in three dimensions. Electrons moving through such an arrangement of nuclei can be described by Bloch functions, a mathematical treatment that shows how the spatial distributions of the electron energies (or momenta) reflect the symmetry of the atomic lattice.

Until now, quasicrystals have defied similar treatment because the rules by which nature arranges their atoms are different from the 'rules of periodicity', even though the positions of their atoms are predictable — that is, quasicrystals are aperiodic but well-ordered. A simple example of this type of order, in one dimension, is the Fibonacci sequence. Each number in the sequence is generated from the sum of the preceding two, and many quasicrystals follow an analogous construction rule in three dimensions to produce long-range atomic order without conventional periodicity. As a result, some quasicrystals have fivefold or tenfold

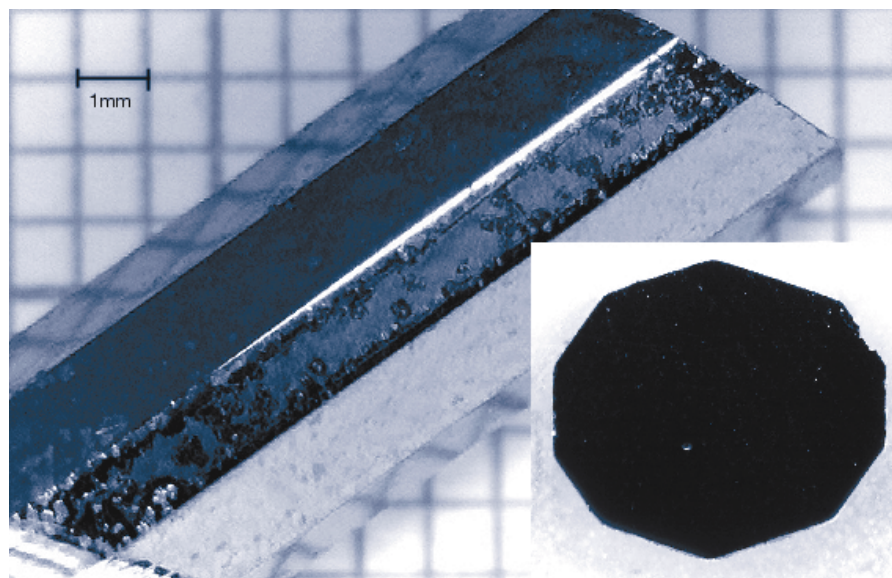


Figure 1 Quasicrystals show unusual symmetry not seen in normal metallic crystals. A single grain⁷ of Al-Ni-Co decagonal quasicrystal shows tenfold symmetry over a millimetre grid. Note reflections of millimetre scale on side facets. Whereas the main figure shows just five of the ten facets clearly, the inset shows a view of a polished surface seen down the tenfold axis with all ten edges visible. The ten facets exposed at the surface reflect the atomic-scale symmetry. Experiments by Rotenberg *et al.*¹ show that electronic states in this quasicrystal also reflect the tenfold symmetry of the crystal structure. Their data also suggest that some of the electrons in a quasicrystal behave like electrons in a normal metal. (Photograph courtesy of I. R. Fisher, N. D. Kelso and P. C. Canfield.)

symmetry (as revealed by their diffraction patterns) — an impossible structure for a truly periodic crystal. How do the electrons, particularly those involved in metallic behaviour, arrange themselves in this quasi-periodic environment?

First, one should note that although quasicrystals are metals, they are not very good metals. This is reflected by their electrical resistivities (resistivity being a measure of the inefficiency of electron transport), which are higher than those of their constituent metallic elements by several orders of magnitude, and much higher than those of conventional metallic alloys. In addition, unlike metals, the electrical resistivity of quasicrystals decreases with increasing temperature and increases with increasing order of the quasiperiodic lattice. This behaviour is clearly correlated with the aperiodic atomic structure, as shown in decagonal quasicrystals with tenfold symmetry (Fig. 1). The decagonal alloys are periodic in one dimension and quasiperiodic in the remaining two. Martin *et al.*² measured the electrical resistivity of a decagonal alloy of Al-Ni-Co and showed that the unusual variation of resistivity with temperature could be measured only within the quasiperiodic plane; electrical transport was normal in the periodic direction.

Rotenberg *et al.*¹ have studied the same Al-Ni-Co alloy as Martin *et al.* to understand better the nature of electronic states in metals with these odd structures. This question has generated much debate, primarily concerning two concepts. First, the density of the most mobile electrons — the most weakly

bound electrons in the electron sea — in a quasicrystal is not as high as in a typical metal. This scarcity of so-called Fermi electrons is called the pseudogap³. Second, the nature of the electronic states may be unique to quasicrystals. Such 'critical states' are in an intermediate state somewhere between localized and delocalized. As a result, the (relatively) few electrons that are energetically available for transport in quasicrystals do not have a good choice of routes, and so constitute a network of lakes and rivers, rather than a sea. (It has not been clear why this effect should be related to the aperiodic structure, but one hypothesis is that large atomic clusters are the key to both⁴.) In fact, it has been suggested that both the pseudogap and the critical states need to be combined to describe successfully the unusual behaviour of electrons in quasicrystals⁵.

Rotenberg *et al.*¹ now provide direct information from angle-resolved photoemission experiments about whether the Fermi electrons in quasicrystals are more localized than in a normal metal. By mapping out the energies and momenta of electrons in the quasiperiodic plane of decagonal Al-Ni-Co, they show convincingly that some of the electronic states must be delocalized and therefore contribute to the electron sea. Furthermore, they show that the electronic states adopt the quasiperiodic symmetry of the atomic cores — a fact that might have been inferred from various previous measurements, including those of Martin *et al.*², but had never been demonstrated explicitly. This shows that a framework



100 YEARS AGO

Mr. Balfour on Scientific Progress.

Phenomena apparently so wide apart as light, radiant heat and electricity, are, as it is unnecessary to remind you, now recognised as substantially identical. From the arrangement of atoms in the molecule, not less than their intrinsic nature, flow the characteristic attributes of the compound... Plausible attempts have been made to reduce the physical universe, with its infinite variety, its glory of colour and of form, its significance and its sublimity, to one homogeneous medium in which there are no distinctions to be discovered but distinction of movement or of stress. And although no such hypothesis can, I suppose, be yet accepted, the gropings of physicists after this, or some other not less audacious unification, must finally, I think, be crowned with success. The change of view which I have endeavoured to indicate is purely scientific, but its consequences cannot be confined to science. How will they manifest themselves in other regions of human activity, in literature, in art, religion? The subject is one rather for the lecturer on the twentieth century rather than for the lecturer on the nineteenth. I, at least, cannot endeavour to grapple with it. From *Nature* 9 August 1900.

50 YEARS AGO

... even the most casual reader of the scientific and technical Press can scarcely fail to be aware of the impatience of the scientific community in both [Great Britain and the United States] at any restraints on freedom of discussion and publication of scientific and technical information which are not plainly justifiable in the public interest. The creation of new barriers, restricting the free flow of information between countries, has not been accepted without protest, and, where accepted, is manifestly viewed with anxiety. Nevertheless, Prof. Bohr's view that the continued secrecy and restrictions deemed necessary for security have themselves split the world community of science into separate camps will not be generally accepted. It may be that a more constructive approach to the problem of the control of atomic energy requires an atmosphere of greater confidence; but Prof. Bohr does not mention the limited and spasmodic participation of Eastern European nations in international scientific conferences. The exchange and interchange of information must be at least bilateral and not unilateral if the advancement of science is to be furthered. From *Nature* 12 August 1950.

analogous to Bloch functions can be applied to quasicrystals.

Rotenberg *et al.* do not, however, rule out the possibility of critical states co-existing alongside the extended states. In this vein, we note that the existence of extended states in decagonal Al-Ni-Co is not too surprising, as it is among the most 'metallic' of the quasicrystals. Icosahedral Al-Pd-Re is at the other extreme, showing extremely high resistivity, and perhaps should even be classified as an insulator rather than as a metal at low temperatures⁶. The next step is to test whether extended states exist in other alloys.

Patricia A. Thiel is at the Ames Laboratory and the

Department of Chemistry, Iowa State University, 1605 Gilman Hall, Ames, Iowa 50011, USA.

e-mail: thiel@ameslab.gov

Jean Marie Dubois is at the Laboratoire de Science et Génie des Matériaux Métalliques (UMR 7584 CNRS-INPL), Ecole des Mines de Nancy, Parc de Saurupt, 54042 Nancy Cedex, France.

e-mail: dubois@mines.u-nancy.fr

1. Rotenberg, E., Theis, W., Horn, K. & Gille, P. *Nature* **406**, 602–605 (2000).
2. Martin, S. *et al.* *Phys. Rev. Lett.* **67**, 719–722 (1991).
3. Belin-Ferre, E. in *Quasicrystals* Vol. 553 (eds Dubois, J.-M. *et al.*) 347 (MRS Conf. Proc., Warrendale, Pennsylvania, 1999).
4. Janot, C. & de Boissieu, M. *Phys. Rev. Lett.* **72**, 1674–1677 (1994).
5. Macia, E. *Phys. Rev. B* **61**, 8771–8777 (2000).
6. Pierce, F. S. *et al.* *Phys. Rev. B* **47**, 5670–5676 (1993).
7. Fisher, I. R. *et al.* *Phil. Mag.* **B79**, 425–434 (1999).

Cancer

Benefits of bad telomeres

Douglas Hanahan

Tumour cells can be considered in terms of their acquisition of capabilities that enable cancerous growth¹. One property involves the ability to go through countless cycles of chromosome replication followed by cell division. This is achieved largely through activation of the enzyme telomerase², which protects the ends (telomeres) of linear chromosomes by capping them with short repetitive DNA elements called telomere repeats. Without telomerase, each chromosome loses a few telomere repeats — and so is shortened — every time it is duplicated. This makes the chromosomes unstable, which can lead to the death of the cell or its inability to continue multiplying. So it is

not surprising that telomerase is activated in most human cancers, often around the stage when progression from contained lesions to invasive cancer occurs³. But the results of Artandi and colleagues⁴, reported on page 641 of this issue, come as a surprise. The authors look at mice engineered to lack functional telomerase and a well-known tumour-suppressor protein, p53. They find that epithelial cancers develop that would not have emerged in mice lacking only p53.

Genome instability can reprogramme a nascent tumour cell, endowing it with the capabilities it needs to develop further¹. The integrity of the genome in normal cells is preserved by a specialized system, a key compo-

nent of which is p53. A signalling circuit centred on p53 detects DNA damage and limits the emergence of cells with mutant genomes, either by arresting the cell-division cycle to allow DNA repair or by inducing cell death⁵. This DNA-damage sensor is often disrupted in tumour cells, allowing genomic instability.

Mice lacking p53 alone rapidly develop sarcomas (tumours of connective tissue) or lymphomas (cancers of white blood cells)⁶. Only rarely are epithelial cancers, such as those of the gut lining, seen in these p53-deficient mice⁶, despite the fact that 90% of human cancers are epithelial in origin, and most show malfunctioning of the p53 damage-sensing pathway^{5,7}. The absence of telomerase alone, however, has no immediate effect on mice⁸, as they have much longer telomeres than humans. But if telomerase-deficient mice are bred with each other, and so too are their progeny, then progressive shortening of the telomeres occurs. After six generations the telomeres are almost non-functional, and the mice begin to show a plethora of abnormalities, including sterility and rare cancers⁸.

Artandi *et al.*⁴ made their unexpected discovery by analysing mice with inactivating mutations in the genes encoding p53 and the essential RNA subunit of telomerase. The authors crossbred these mice, and again found that rapidly developing sarcomas and lymphomas were prevalent in the offspring. Reasoning that the early onset of such tumours might be masking other, more slowly developing cancers, the authors sought to limit the sarcomas and lymphomas, by using telomerase-deficient mice that carried one normal and one mutant p53 gene (that is, they were 'heterozygous' for p53). Mice heterozygous for p53 (but with normal telomerase) develop sarcomas and lymphomas later than mice with two mutant versions of the p53 gene⁶. This probably reflects incomplete tumour suppression by the single intact copy of p53 and the relative ease with which it can be lost in tumours.

Artandi *et al.* inbred their telomerase-deficient, p53-heterozygous mice for five to seven generations. The resulting mice had short telomeres. As they aged, about half of the mice still developed sarcomas and lymphomas. But the remainder developed epithelial carcinomas, mainly of the breast, skin or intestines. Intriguingly, all the mice analysed had microscopic cancerous lesions in the large intestine, indicating a marked effect on this self-renewing epithelium. What caused the shift in tumour spectrum?

The data indicate⁴ that the already short telomeres may have become critically short in these epithelial tissues, presumably during cell proliferation, with the malfunctioning telomeres catalysing genomic rearrangements that initiated and sustained the development of cancer. The tumours showed marked disarray in their genomes, presum-

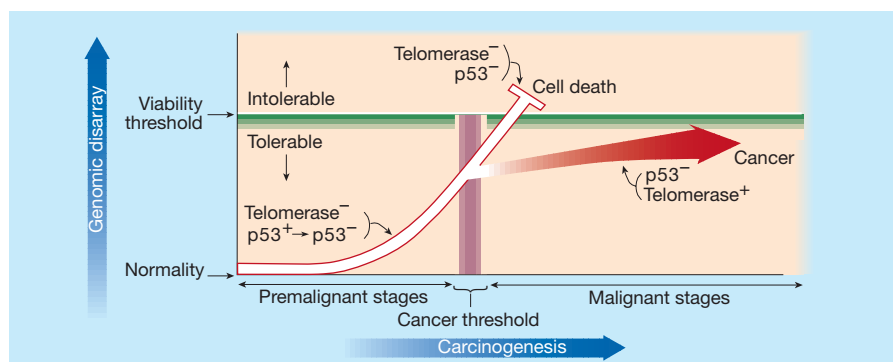


Figure 1 How the absence and presence of the enzyme telomerase might benefit human cancer cells. This model is suggested by both present⁴ and previous^{2,3,10,11} results. Cancer development (carcinogenesis) occurs in several stages, and can be considered in terms of accumulating disarray of the genome, produced by genomic instability. There are two thresholds: a viability threshold, above which genomic disarray is incompatible with the survival of the cell; and a cancer threshold, after which developing tumours grow expansively. In the early (premalignant) stages, continuing cell proliferation in the absence of telomerase results in progressive shortening of telomeres, which eventually become so short that they can no longer protect chromosome ends. This causes chromosome fusions, translocations and other rearrangements. Combined with the loss of DNA-damage sensors (such as p53), the scrambled genomes persist and acquire mutations that enhance the capabilities of the tumour. If unchecked, the accumulating genomic disarray will reach intolerable proportions, resulting in cell death. But in many developing cancers, telomerase is activated around the time of the cancer threshold, restoring telomere function, reducing further genomic instability and allowing expansive proliferation.

ably allowed by the observed loss of the one functional p53 gene. The disarray included chromosomal end-to-end fusions and translocations, as well as changes in chromosome number. It seems that these genomic rearrangements allowed the development of epithelial cancers in particular. But how can one relate this result — that the absence of telomerase can lead to epithelial cancers in mice — with extensive evidence that telomerase activity is common to, and important for, those very epithelial cancers in humans? The answer may lie in the timing of telomerase activation, and in the nature of cancer progression (Fig. 1).

Cancers develop in several distinct steps, with cells of increasing aggressiveness and capability emerging over time. In most cancers, disease progression is accompanied by an accumulation of genomic abnormalities⁹; in some, the disarray arises mainly at a characteristic stage. In the absence of telomerase, as telomeres become too short to function, sudden episodes of widespread genomic disarray can occur — a conclusion supported by previous work^{10,11}. This might enable cells to acquire the capabilities they need to become tumour cells. But genome disarray can reach a point (a 'viability threshold') beyond which cells can no longer live. So the eventual activation of telomerase could be seen as necessary for the survival of a human tumour cell (Fig. 1). Early on, the normal absence of telomerase allows critical telomere shortening in the proliferating cells. This shortening, combined with defects in a DNA-damage sensor (such as p53), produces genome rearrangements that allow the cancer to develop. But then, before the damage becomes so extensive as to preclude cell viability, telomerase is activated, adding telomere repeats and stabilizing the ends of both normal and mutant chromosomes. This enables continuing proliferation, progression and dissemination of the cancer.

One testable prediction of this hypothesis is that the genome rearrangements produced by malfunctioning telomeres alter the expression or function of critical genes, allowing the genesis of epithelial cancers. High-resolution genome mapping and studying candidate genes would address this possibility. But why do the genome rearrangements produced by telomere malfunction fuel epithelial cancers specifically? Perhaps the growth-regulatory genes in epithelial cells are, coincidentally, located in regions prone to being altered by genome rearrangements induced by malfunctioning telomeres. Or perhaps the regulation of transitory proliferation in self-renewing epithelial sheets is more easily disrupted by chromosomal imbalances. The answer may shed light on the bias towards epithelial cancer in humans.

A challenge to the theory comes, however, from the observation that the mouse tumours seem to violate the viability thresh-

old by their very emergence in the telomerase-deficient mice. This result might be explained by the presence in mice of an alternative, DNA-recombination-based mechanism of preserving telomeres¹². Also, when two chromosomes are fused end to end, burying malfunctioning telomeres, the hybrid chromosomes in one orientation seem to be better tolerated in mice than in humans.

The theory (Fig. 1) does fit, though, with the patterns of telomerase activity seen in many human cancer types³. If progression through the early stages of cancer is indeed fuelled in part by the absence of telomerase, there may be implications for treatments that seek to interfere with telomerase activity. Pharmacological inhibitors of telomerase are likely to cause advanced cancers to breach the viability threshold and die. But they may also enhance the progression of telomerase-expressing cancers that are still far from the threshold. So telomerase inhibitors may

prove most effective when used in combination with traditional chemotherapeutic drugs that damage DNA, together accelerating the accumulation of intolerable disarray in the genome of a cancer cell. ■

Douglas Hanahan is in the Department of Biochemistry and Biophysics, Hormone Research Institute, and the Comprehensive Cancer Center, University of California at San Francisco, San Francisco, California 94143, USA.
e-mail: dh@biochem.ucsf.edu

1. Hanahan, D. & Weinberg, R. *Cell* **100**, 57–70 (2000).
2. Harley, C. B. *et al. Cold Spring Harb. Symp. Quant. Biol.* **59**, 307–315 (1994).
3. Shay, J. W. & Bacchetti, S. *Eur. J. Cancer* **33**, 787–791 (1997).
4. Artandi, S. E. *et al. Nature* **406**, 641–645 (2000).
5. Levine, A. J. *Cell* **88**, 323–331 (1997).
6. Attardi, L. D. & Jacks, T. *Cell. Mol. Life Sci.* **55**, 48–63 (1999).
7. Sherr, C. J. *Genes Dev.* **12**, 2984–2991 (1998).
8. Rudolph, K. L. *et al. Cell* **96**, 701–712 (1999).
9. Gray, J. W. & Collins, C. *Carcinogenesis* **2**, 443–452 (2000).
10. de Lange, T. & Jacks, T. *Cell* **98**, 273–275 (1999).
11. Artandi, S. E. & DePinho, R. A. *Nature Med.* **6**, 852–855 (2000).
12. Kass-Eisler, A. & Greider, C. W. *Trends Biochem. Sci.* **25**, 200–204 (2000).

Oceanography

The giant diatom dump

Victor Smetacek

Diatoms, single-celled algae with silica cell walls, are a major component of oceanic phytoplankton. They are a motley lot. There are about 1,000 species, ranging across three orders of magnitude in size (about as much as land plants) and exhibiting a remarkable variety of shapes. The biggest species take the form of flat discs, squat cylinders or slender rods, and can reach 2–5 millimetres in diameter or length.

These giants among diatoms occur in all of the oceans and are prominent in the sediments on the ocean floor. Their ecology is not well understood, but they are widely believed to have only a marginal influence in marine ecosystems. In a paper in *Deep-Sea Research II*, however, Kemp *et al.*¹ show that giant diatoms can make a large contribution to the rain of particles settling out of the ocean surface layer, and so are serious players in ocean biogeochemistry.

In contrast, the small-celled, long-chained diatoms, which form dense blooms along ocean margins and across the North Atlantic, are well studied. They occur in spring after deep winter mixing introduces new nutrients to the surface layer, and also in freshly upwelled, nutrient-rich water. Given their high growth rates, these small-celled species (5–50 micrometres) accumulate biomass rapidly following warming and stabilization of a shallow, nutrient-rich surface layer. They tend to aggregate into flocs that sink quickly when nutrients run out. Particles sinking in the aftermath of these blooms fuel the 'biological carbon pump' that draws down carbon

dioxide from the atmosphere and exports it to the ocean interior². Kemp *et al.* now report that the springtime and upwelling blooms leave less of a trace in the sediments than the large diatoms that settle out later in the year.

This new perspective comes from detailed analysis of the annual sequence of settling particles preserved in layered sediments from sites in the Gulf of California and the Mediterranean. Such layered sediments are rare in the ocean (in contrast to lakes) because the burrowing activity of the deep-sea fauna tends to blend seasonal signals over many years. In the sediments painstakingly examined by Kemp *et al.*, however, discrete bands representing seasonal settling events have been preserved. Layers of large diatoms were prominent at both sites. The layers were dominated by single species (Fig. 1), which in each case had settled during the summer and autumn months. Kemp *et al.* argue convincingly that these layers reflect a different type of event to that producing the blooms. They propose that the layers are the outcome of a 'fall dump' of large diatoms, as opposed to that of the 'spring bloom' of their smaller relatives.

The prominence of the fall dumps is surprising because the large-celled diatoms must have accumulated substantial biomass in stratified, nutrient-depleted water columns. Stratification arises when the surface layer warms, becomes less dense and is thus sealed off from the rest of the upper ocean by the thermocline, a barrier created by an abrupt change in temperature. Under

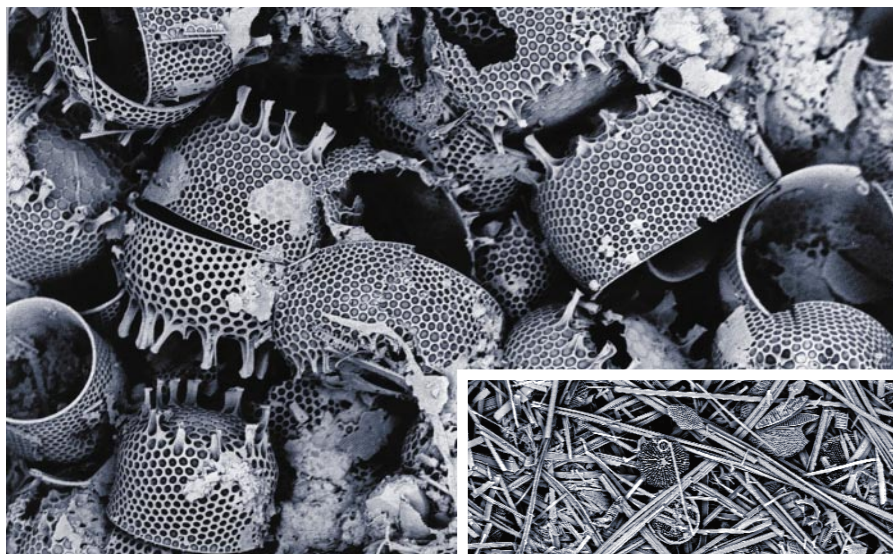


Figure 1 Evidence of the 'fall dump' of giant diatoms identified by Kemp *et al.*¹. Above, a sediment layer from the Gulf of California dominated by cell walls of *Stephanopyxis palmeriana* (diameter < 70 μm). Right, a sediment layer from the Southern Ocean, showing a tangled mass of the rod-shaped species *Thalassiothrix antarctica*, which grows up to 4 mm in length. (Courtesy of A. E. S. Kemp.)

such conditions, which are characteristic of most of the surface ocean, regeneration of nutrients by microbes drives phytoplankton production, which is dominated by the smallest (up to 5 micrometres) size classes. This is because surface-to-volume ratio increases as size decreases, improving the efficiency of nutrient uptake and hence growth rate. So how do the slow-growing giant diatoms acquire their nutrients?

Large diatoms can tap into different nutrient pools by harbouring other, nitrogen-fixing organisms³, or by using buoyancy regulation to migrate vertically between the surface and the deep, nutrient-rich water below the thermocline⁴. But many species appear to do neither. Kemp *et al.* suggest that some of the more prominent species are members of a 'shade flora' that grow slowly in the less well lit waters at the base of the surface layer, taking advantage of episodic nutrient input when the thermocline is eroded (by passing storms, for instance)⁵. In these circumstances, however, large diatoms would still be out-competed by small ones.

But large diatoms do have advantages over smaller cells. First, their lower abundance reduces the rate at which they encounter pathogens and predators. Second, their size, and their especially robust silica cell wall, make them tough nuts for predators adapted to crop the more abundant but less well armoured smaller cells (which are eaten almost as fast as they are produced). Relative investment in mechanical defence declines with decreasing surface-to-volume ratio, so large diatoms can accumulate in biomass

because of their low mortality rate — a point borne out by observations that their populations tend to persist for many weeks¹. By simply living longer they can slowly sequester nutrients from the rapidly spinning microbial loop. Such species are the thistles of the plankton meadows — an end stage in the seasonal cycle of a heavily grazed plant community.

Because the giant diatoms are built for durability, it follows that a greater proportion of their cell walls reach and are buried in the sediments than are those of the weakly silicified, small-celled diatoms. Most of the particles sinking out of the surface layer are consumed by other organisms within the upper

200 metres. But the remains of the armoured giants traverse this zone with smaller losses than any other phytoplankton cells, helped on the way by Stokes's law: the bigger they are, the harder they fall. Data from sediment traps (moored funnels that collect particles sinking through the water column at pre-set intervals) support this view^{6,7}. So their remains are more prominent in the annual cycles of the sediments examined by Kemp *et al.* than are those of the spring blooms. Big diatoms are deep exporters.

The paper by Kemp *et al.* raises questions for ecologists, such as how form is shaped by function in phytoplankton. Beyond that, it is a showcase example of how the gap between biological oceanography and the new science of palaeoceanography can be bridged. These disciplines have been kept apart by the different timescales they address, and also by separate funding agencies and pools of reviewers. The palaeoceanographers need help in deciphering the vast storehouse of blurred information buried in the sedimentary documents. Giant diatoms are an especially worthwhile target for study, because they include species adapted to specific environmental conditions and so can be precise indicators of past climatic regimes. Improving understanding of their ecology, and so their use as climate proxies, will require fresh thinking and interdisciplinary collaboration both during data collection at sea and modelling at the computer.

Victor Smetacek is at the Alfred-Wegener-Institute for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany.

e-mail: vsmetacek@awi-bremerhaven.de

1. Kemp, A. E. S., Pike, J., Pearce, R. B. & Lange, C. B. *Deep-Sea Res.* **47**, 2129–2154 (2000).
2. Smetacek, V. *Protist* **150**, 25–32 (1999).
3. Villareal, T. A., Altabet, M. A. & Culver-Rymysza, K. *Mar. Ecol. Prog. Ser.* **76**, 201–204 (1991).
4. Villareal, T. A. *et al.* *Nature* **363**, 709–712 (1993).
5. Goldman, J. C. *Deep-Sea Res.* **140**, 159–168 (1993).
6. Takahashi, K. *Deep-Sea Res.* **33**, 1225–1251 (1986).
7. Scharek, R. *et al.* *Mar. Ecol. Prog. Ser.* **182**, 55–67 (1999).

Cell biology

Making membranes in bacteria

Rosemary A. Stuart and Walter Neupert

Cellular membranes are formed by the insertion of newly synthesized components, such as lipids and proteins, into pre-existing membranes. Each membrane is made up of a specific and unique set of proteins, which determines its identity. As a consequence, the insertion of proteins cannot be a spontaneous event. Rather, it relies on a set of 'protein translocases', which sit in membranes and move newly made proteins into or through them (Fig. 1, overleaf). In Gram-negative bacteria, which are bounded by two membranes, six translocases have been

found so far — two in the inner and four in the outer membrane. In eukaryotic cells (those with a nucleus), about ten different protein translocases have been identified, perched in the membranes of the different organelles such as the endoplasmic reticulum, mitochondria and chloroplasts. One might think this would be plenty.

But, on page 637 of this issue¹, Samuelson and colleagues describe yet another translocase, called YidC, which inserts proteins into the inner membrane of bacteria. Their results provide further support for a bacteri-

al origin of mitochondria and chloroplasts.

Bacteria use translocases not only to insert proteins into membranes, but also to move proteins through the membranes to the outside of the cell. Most bacterial proteins secreted in this way use a general translocase called SecYEG to move across the inner (plasma) membrane^{2,3}. This type of movement is said to be Sec-dependent. Another translocase, the Tat system, exports several proteins that contain firmly bound co-factors into the space between the inner and outer bacterial membranes⁴. Many proteins that remain in the plasma membrane are also inserted in a Sec-dependent manner (Fig. 1).

But the discovery of the new translocase was not entirely unexpected. First, not all inner-membrane proteins in bacteria use SecYEG for insertion, so another translocase was known to be needed. Second, a protein discovered in the inner membrane of mitochondria, the powerhouses of eukaryotic cells, is required for the formation of complexes of membrane proteins^{5,6}. This protein, called Oxa1, has a similar amino-acid sequence to YidC, a protein of unknown function in the bacterium *Escherichia coli*. It now appears¹ that YidC is the missing bacterial translocase, or at least a component of it. YidC is located in the plasma membrane in both Gram-negative and Gram-positive bacteria (which have no outer membrane). It also seems to be evolutionarily conserved: it is akin not only to Oxa1 in mitochondria^{5,6} but also to the translocase Albino3 in chloroplasts⁷.

The story starts with studies of protein insertion into the inner mitochondrial membrane in yeast. Several mitochondrial proteins (encoded in the nuclear or the mitochondrial genome) are inserted into this inner membrane from inside the mitochondrion. This process was known to occur in a Sec-independent manner^{8–10}, as yeast mitochondria do not contain homologues of the bacterial SecYEG machinery. Oxa1 was then shown to be the translocase involved^{8–10} and to interact directly with proteins as they are inserted into the inner membrane¹⁰. In terms of energetic requirements and adherence to the 'positive inside' rule of protein sorting, the pathway taken by these mitochondrial proteins resembled the Sec-independent pathway used by bacterial membrane proteins^{8–11}. So YidC was proposed to have a similar function in bacterial plasma membranes to that of Oxa1 in mitochondrial inner membranes^{8,10}.

Samuelson *et al.*¹ now confirm this suspicion. They study the Procoat protein from the virus M13. This protein uses a Sec-independent pathway for insertion into the bacterial inner membrane. For some time, it was believed to insert spontaneously into the membrane¹². But, when Samuelson *et al.* reduced the expression — and hence the steady-state levels — of YidC in *E. coli* cells, they observed that the insertion of newly produced Procoat protein was almost com-

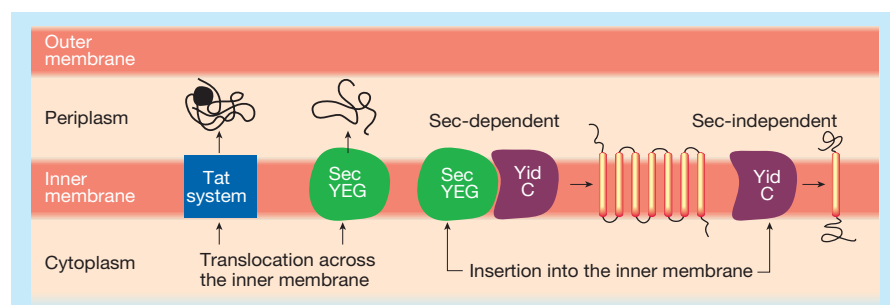


Figure 1 Pathways of protein translocation across, and insertion into, the bacterial inner membrane. The Tat system is used to move proteins containing bound co-factors (black shape) across the inner membrane into the periplasm. The SecYEG system is used to move other proteins into the periplasm, where some of them will stay; others will be transported further, either into or across the outer membrane. SecYEG is also required for the insertion of proteins into the inner membrane. Samuelson *et al.*¹ have now identified YidC as another 'translocase' in the bacterial inner membrane. It may work either independently or together with SecYEG to insert proteins into the inner membrane, depending on the protein being inserted.

pletely inhibited. Samuelson *et al.* also show that other Sec-independent membrane proteins depend on the presence of YidC for insertion, and that such proteins contact YidC directly. So the Sec-independent insertion of proteins into bacterial membranes does not occur spontaneously, but rather is facilitated by YidC.

The authors also found that the membrane insertion of a Sec-dependent protein was affected to some extent by the decrease in YidC expression. This fits with the finding¹³ that YidC associates with the SecYEG translocase. And membrane proteins such as FstQ and leader peptidase, the insertion of which into the inner membrane is Sec-dependent, associate temporarily with YidC^{13,14}. All these findings suggest that, for some proteins, the two pathways for insertion into the bacterial plasma membrane converge. A model for the reliance of the Sec-dependent pathway on YidC might include the interaction of a growing polypeptide with SecYEG. YidC would then constitute another component of the translocation machinery, probably with the specialized function of mediating the insertion of membrane-spanning segments of the newly formed protein^{1,13} (Fig. 1).

These results are also complemented by the finding that Albino3 is required for the Sec-independent insertion of the light-harvesting chlorophyll-binding protein into chloroplast membranes¹⁵. So, YidC in bacteria, Albino3 in chloroplasts, and Oxa1 in mitochondria have similar functions, showing that the Sec-independent pathway of protein insertion into membranes has been conserved throughout evolution. This fits with the 'endosymbiont hypothesis' for the origin of mitochondria and chloroplasts. This hypothesis proposes that the inner membrane of mitochondria and the thylakoid membrane of chloroplasts are derived from the plasma membranes of their respective bacterial ancestors.

We now need to know exactly how members of the YidC/Oxa1/Albino3 protein fam-

ily mediate the insertion of proteins into membranes. How are proteins that use this translocase recognized? How and when are hydrophobic membrane-spanning segments released into the lipid phase of the membrane? How does YidC interact with the SecYEG complex, in physical and functional terms? The Sec61 complex of the endoplasmic reticulum is derived from the bacterial SecYEG machinery. Does YidC have a similar function to the TRAM protein, which cooperates with the Sec61 complex to insert proteins into membranes¹⁶? Other components must assist YidC/Oxa1/Albino3, but what are they? Given that Oxa1 can function without a Sec-type translocase, can YidC and Albino3 also work on their own? An arsenal of cell-biological techniques — including the purification and functional reconstitution of the members of the protein family — will be needed to answer these questions. With the results should come a more complete understanding of how membranes are put together. ■

Rosemary A. Stuart is in the Biology Department, Marquette University, 530 North 15th Street, Milwaukee, Wisconsin 53233, USA.

e-mail: rosemary.stuart@marquette.edu

Walter Neupert is at the Institut für Physiologische Chemie der Universität München, Goethestrasse 33, D-80336 München, Germany.

e-mail: neupert@bio.med.uni-muenchen.de

- Samuelson, J. C. *et al.* *Nature* **406**, 637–641 (2000).
- Wickner, W. & Leonard, M. R. *J. Biol. Chem.* **271**, 29514–29516 (1996).
- Rapoport, T. A. *et al.* *Annu. Rev. Biochem.* **65**, 271–303 (1996).
- Dalbey, R. E. & Robinson, C. *Trends Biochem. Sci.* **24**, 17–21 (1999).
- Bauer, M. *et al.* *Mol. Gen. Genet.* **245**, 272–278 (1994).
- Bonnefoy, N. *et al.* *J. Mol. Biol.* **239**, 201–212 (1994).
- Sundberg, E. *et al.* *Plant Cell* **9**, 717–730 (1997).
- He, S. & Fox, T. D. *Mol. Biol. Cell* **8**, 1449–1460 (1997).
- Hell, K. *et al.* *FEBS Lett.* **418**, 367–370 (1997).
- Hell, K. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 2250–2255 (1998).
- Rojo, E. E. *et al.* *J. Biol. Chem.* **274**, 19617–19622 (1999).
- Geller, B. L. & Wickner, W. *J. Biol. Chem.* **260**, 13281–13285 (1985).
- Scotti, P. A. *et al.* *EMBO J.* **19**, 542–549 (2000).
- Houben, E. N. G. *et al.* *FEBS Lett.* (in the press).
- Moore, M. *et al.* *J. Biol. Chem.* **275**, 1529–1532 (2000).
- Johnson, A. E. & van Waas, M. A. *Annu. Rev. Cell Dev. Biol.* **15**, 799–842 (1999).

Seafloor spreading

Portrait of a magma chamber

Robert S. Detrick

About 20 km³ of new ocean crust is formed each year along the spreading centres of the global mid-ocean-ridge system. These spreading centres are usually not simple, linear structures, but are broken into offset and sometimes overlapping segments. Most of the molten rock (melt or magma) that solidifies into ocean crust resides within shallow magma chambers beneath the axes of the ridges before erupting at the seafloor or 'freezing' onto the edges of the separating tectonic plates.

Not surprisingly, the relationship between crustal magma chambers, the segmentation of spreading centres, and the supply of melt from the underlying mantle is still poorly understood. On page 614 of this issue, in a paper that constitutes an important step towards improving that understanding, Kent and co-workers¹ present the first three-dimensional seismic reflection images of a

mid-ocean-ridge magma system. Their study focuses on the East Pacific Rise near 9° 03' N, about 500 km off the coast of Mexico.

Two different models have been proposed to explain how melt is supplied to the crust along fast-opening spreading centres such as the East Pacific Rise. In one, mantle upwelling is two-dimensional and sheet-like, and melt is fed into crustal magma chambers along the entire length of each spreading segment². In the other, known as the focused magma supply model, mantle upwelling is plume-like and three-dimensional, with widely spaced centres of magma injection into the crust and lateral transport of melt in the crust towards the ends of ridge segments^{3,4}.

Some of the systematic variations in seafloor depth along the East Pacific Rise axis appear to support the focused magma supply model³. Such variations include segment centres that are inflated and shallow, and a deepening of the ridge axes towards segment ends. However, seismic reflection and crustal tomography studies⁵⁻⁷ in these same areas have found that crustal magma bodies occur roughly every 10–15 km, which apparently

limits the scale of magma mixing along the ridge axis to distances much smaller than a segment half-length (about 20–50 km). This picture is complicated further by a few reflection profiles⁸⁻¹⁰ across overlapping spreading centres, which show that, rather than being magma-starved, there is melt and normal or even anomalously thick crust at some of the offsets between segments.

Overlapping spreading centres occur where the 'neovolcanic zone', the narrow band of most recent volcanism, is offset by a short distance (0.5–10 km) and two adjacent neovolcanic zones overlap and bend around a small basin located at the offset³ (Fig. 1). Larger overlapping centres mark the ends of morphologically defined segments and so provide a unique place to test competing models of magma supply.

Fully molten magma has a seismic velocity that is less than half that of solid rock. This large contrast in velocity gives rise to strong reflections from within the oceanic crust that can be used to map the presence and geometry of magma bodies. By collecting 200 profiles, each separated by just 100 m, Kent *et al.* have constructed a fully three-dimensional image of crustal reflectivity in a 400 km² area encompassing nearly the entire overlapping spreading centre on the East Pacific Rise at 9° 03' N, and have mapped the location of magma bodies throughout this area.

They find narrow, surprisingly continu-

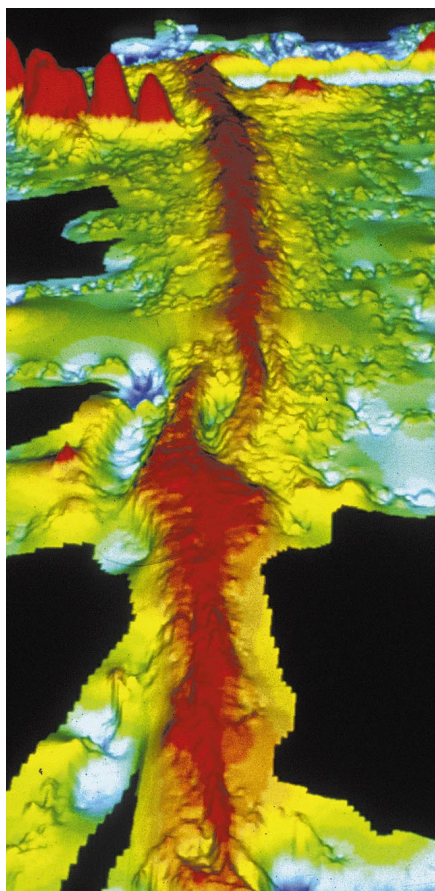


Figure 1 View looking north along the East Pacific Rise. The spreading axis is offset by about 10 km near 9° 03' N (centre of image). Here, the 'neovolcanic zone' — the narrow band of most recent volcanism — overlaps and curves around a small basin located at the offset.

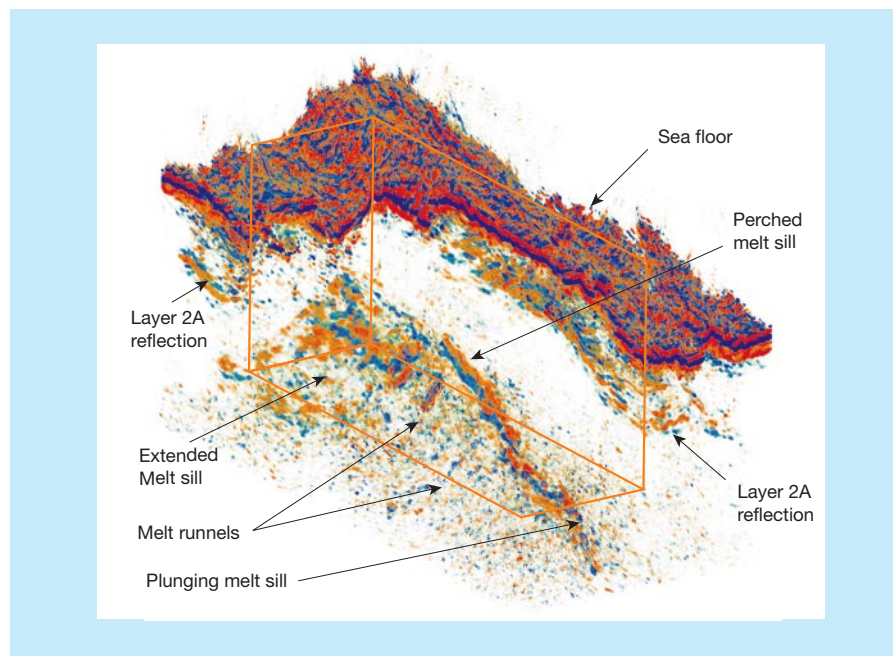


Figure 2 Mid-ocean-ridge magma chambers in three dimensions. This is a view, looking northeast, of seismic reflectivity beneath the eastern limb of the 9° 03' N overlapping spreading centre on the East Pacific Rise. The red and blue colours on the upper surface are reflections from the sea floor; the orange and blue-green hues on the lower surface are reflections from melt bodies within the oceanic crust. Magma present beneath the overlap basin (on the left) appears to be migrating upwards (to the right) where it ponds in a narrow, perched melt sill beneath the eastern ridge axis. These results suggest that vertical melt transport from the underlying mantle, together with subsequent focusing of melt beneath a magma 'freezing' boundary in the mid-crust, supply melt to the ridge axis in this area. (Reproduced from ref. 1.)

ous 'lenses' or 'sills' of melt beneath both limbs of the overlap, but no large melt body beneath the overlap basin (although there is a crustal low-velocity zone under the basin¹¹, which may indicate the presence of small pockets of melt). Although the geometry of the melt lens underlying the western limb of the overlap is relatively simple, that associated with the eastern — propagating — limb is fascinatingly complex (Fig. 2).

At its southern end the melt sill is only a few hundred metres wide ('plunging melt sill' in Fig. 2). It falls some 500 m in depth relative to the sea floor, displaying several small (200 m) right-lateral steps within the zone of deepening. To the north, the melt horizon is displaced west of the axis of the ridge and is some 4 km wide ('extended melt sill' in Fig. 2). The melt body in this area appears to occur at about a constant depth below the base of a shallower reflector (the layer 2A reflection), which is often interpreted as the contact between lavas and an underlying layer of quickly cooled, melt-filled cracks called dikes that feed these eruptions¹². Kent *et al.* suggest that buoyant melt rising vertically from below gets trapped beneath the dike section, which may be impermeable to the melt, then flows 'uphill' in rivulets or sheets along the base of the sheeted dike complex towards the ridge axis in a process analogous to the way oil and gas migrate upwards and become trapped in sedimentary rocks.

Kent and co-workers believe that a combination of vertical melt transport from the underlying mantle, and subsequent focusing of melt beneath a magma freezing boundary in the mid-crust, supplies melt to the ridge

axes in this area. So they think that the ridge axes are not being supplied by magma transported laterally from the distant segment centre, as predicted by the focused magma supply model. Their observations cannot rule out focused flow in the mantle, with melt migration along the ridge axis beneath the crust. But from the geometry of the melt sill observed at the 9° 03' N segment overlap, it seems that crustal transport of magma from the segment centre cannot be the main source of melt to the ridge axis in this region.

As the first three-dimensional reflection experiment on a mid-ocean ridge, this study is a milestone. In the future, three-dimensional imaging of the structure of the crust, as well as of the underlying mantle, will be increasingly seen as essential for understanding the volcanic, tectonic and hydrothermal processes that accompany the formation of oceanic crust.

Robert S. Detrick is in the Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA. e-mail: rdetrick@whoi.edu

1. Kent, G. M. *et al.* *Nature* **406**, 614–618 (2000).
2. Parmentier, M. & Phipps Morgan, J. *Nature* **348**, 325–328 (1990).
3. Macdonald, K. *et al.* *Nature* **335**, 217–225 (1988).
4. Wang, X. & Cochran, J. J. *Geophys. Res.* **98**, 19505–19531 (1993).
5. Toomey, D., Purdy, G., Solomon, S. & Wilcock, W. *Nature* **347**, 639–645 (1990).
6. Dunn, R. & Toomey, D. *Nature* **388**, 259–262 (1997).
7. Singh, S. *et al.* *Nature* **394**, 874–878 (1998).
8. Mutter, J. *et al.* *Nature* **336**, 156–158 (1998).
9. Collier, J. & Sinha, M. *Nature* **346**, 646–648 (1990).
10. Kent, G., Harding, A. & Orcutt, J. J. *Geophys. Res.* **98**, 13971–13996 (1993).
11. Bazin, S. K. *Three-dimensional Crustal Structure of East Pacific Rise Discontinuities from Tomographic Inversions*. Thesis, Univ. California, San Diego (2000).
12. Harding, A., Kent, G. & Orcutt, J. J. *Geophys. Res.* **98**, 13925–13944 (1993).

Chromatin regulation

Formatting genetic text

Renato Paro

Writing words in bold or italics does not change the overall meaning of a passage of text. But it does emphasize the importance of that part of the text, changing the way in which it should be read. One could compare the formatting of text in this way to the effects of modifying histone proteins on the genetic information stored in DNA sequences. Histone proteins have a structural role in the nucleus: they bundle up DNA and other proteins, producing a compact structure called chromatin. They also have regulatory roles in various nuclear processes, such as the control of the transcription of genes into RNA copies. The involvement of histones in these processes is itself regulated by various modifications of the histone amino-terminal tails. The idea of a combinatorial 'code' of histone modifications has been proposed to complement the

information stored in DNA sequences^{1,2}, in much the same way that highlighting written words in bold or italics complements the information that they carry.

On page 593 of this issue³, Rea *et al.* describe a new way of formatting genetic text. They show that a mammalian protein found in chromatin can add methyl groups to specific sites in the histone protein H3, and that this type of formatting is required for the cell to proceed through nuclear division.

Several enzymes can modify histone tails. For example, histone acetyltransferases transfer acetyl groups to the histone tails, and kinases add phosphate groups. In most cases, these enzymes seem to be able to act on a broad range of histone proteins that are associated with a variety of nuclear processes. Rea *et al.*³ have now identified a methyltransferase that appears to act almost

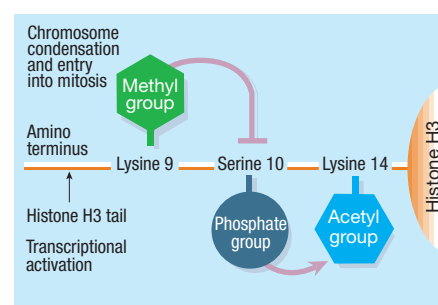


Figure 1 The genetic equivalent of formatting text. Histone proteins bundle up DNA and other proteins into a structure called chromatin. The amino-terminal tails of histones can be modified. Rea *et al.*³ identify an enzyme that modifies histone H3 by adding methyl groups, and show that different modifications of H3 may be interdependent. Methylation at the amino acid lysine at position 9 in the histone H3 tail inhibits the addition of a phosphate group to serine 10 by the enzyme Ipl1/aurora. This phosphorylation event is needed for the correct entry of cells into mitosis. Phosphorylation of serine 10 helps the histone acetyltransferase GCN5 to acetylate lysine 14. These events are correlated with the activation of transcription of a particular set of yeast genes⁶ and of mitogen-induced immediate-early genes in mammalian cells⁷.

exclusively at specific sites in histone H3. By showing that this particular modification of histone H3 influences progression through mitosis, the authors are also able to establish a connection between a defined histone modification and a specific nuclear process.

Rea *et al.* look at two genes, human *SUV39H1* and mouse *Suv39h1*, which influence the formation of heterochromatin (the part of the genome that is never transcribed) and the segregation of chromosomes during nuclear division (mitosis). This family of chromatin proteins is characterized by the presence of two evolutionarily conserved motifs — strings of particular amino acids in a specific order — called the chromodomain and the SET domain⁴. Using *in vitro* assays, the authors show that it is the SET domain that has histone methyltransferase activity, covalently attaching a methyl group to the amino acid lysine at position 9 in the tail of histone H3.

Surprisingly, when Rea *et al.* tested a variety of other proteins that contain a SET domain, they found that not all of these proteins are histone methyltransferases. For example, proteins of the so-called Enhancer-of-Zeste or Trithorax classes — the founding members of the 'SET-domain' family⁴ — do not add methyl groups to histone H3. The authors resolve this discrepancy by showing that cysteine-rich regions next to the SET domain are also needed for the histone methyltransferase activity. Only proteins of the SUV39H class contain all the necessary regions.

But finding out that this extended SET domain has methyltransferase activity *in vitro*

was the (relatively) easy part. How does one show that histone H3 is also the substrate of the SUV39H proteins in living cells? The histone-encoding genes are repeated in the genome of most eukaryotic cells. This makes it hard to alter all the histone H3 genes so as to observe what happens when the protein target of the methyltransferase activity is manipulated. However, Rea *et al.* find that the methylation of lysine 9 influences subsequent modifications of adjacent amino acids. Phosphorylation of serine residue 10 on histone H3 is important during mitosis and for the activation of transcription. The mitotic kinase Ipl1/aurora phosphorylates serine 10 in histone H3 (ref. 5). This modification inhibits *in vitro* methylation of lysine 9. Similarly, methylation of lysine 9 appears to inhibit the phosphorylation of serine 10 (Fig. 1), so the modification signals may be interdependent.

These observations allow Rea *et al.* to get round the problem of the presence of several histone genes in the mouse genome. They show that this crosstalk is also observed in fibroblast cells taken from mice in which the function of the two known *Suv39h* genes has been knocked out. A significant proportion of these mutant fibroblasts have aberrant nuclear morphologies typical of mitotic defects. More important in this context, the level of phosphorylation of serine 10 is higher in these cells. So, because lysine 9 cannot be fully methylated in these cells, the phosphorylation of serine 10 is inhibited less, apparently resulting in incorrect regulation of mitotic events. Another regulatory connection — a synergism between phosphorylation of serine 10 and acetylation of lysine 14 in histone H3 (refs 6, 7) — has been correlated with events in transcriptional regulation (Fig. 1). This interdependence of tail modifications argues in favour of a role for a 'histone code' in regulating nuclear processes.

So it seems that complex, and sometimes interdependent, patterns of histone-tail modifications can be associated with particular chromosome regions and functions. We are starting to identify the enzymatic entities that establish these patterns. One of these, the extended SET domain, has now been shown to be a methyltransferase that is specific to histone H3. The list of enzymatic protein domains will almost certainly expand rapidly. It is interesting that many chromatin proteins, apparently with wildly differing functions, share different combinations of such conserved protein motifs. These proteins may use a combinatorial set of enzymatic functions to modify histones differently, so that the histones can in turn control a range of nuclear functions.

The challenge for the future is to define how this type of formatting is read and interpreted. Do histone-tail modifications have merely structural effects, resulting in changes in the organization of chromatin? Might they thereby lead to open or closed

chromatin conformations, and so regulate the accessibility of the primary information — that encoded in DNA? Or do histone patterns themselves carry information, having a direct effect on a nuclear process and eventually becoming heritable like the genetic information? Either way, it is clear that reading mechanisms must exist that can distinguish a bold from an italicized section of the genetic text. ■

Superconductivity

On the verge of magnetism

Piers Coleman

During the past two decades, condensed-matter physicists have had to adjust to a climate of discovery in which many of the long-accepted rules of solid-state physics are boldly flouted by materials with unexpected properties. These 'emergent properties', such as high-temperature superconductivity, quantized Hall conductivity and giant magnetoresistance, arise in complex materials in which electrons develop different states of organization and correlation. For example, physicists have long thought of magnetism and superconductivity as being like oil and water. One of the hallmarks of a superconductor is the expulsion of a magnetic field passing through it. Likewise, tiny amounts of magnetic impurities are usually enough to completely eliminate superconductivity. But in a new twist, Saxena *et al.*¹ report on page 587 of this issue the discovery of bulk superconductivity in the harshest possible environment: inside a metallic ferromagnet.

A metallic ferromagnet is a permanent magnet such as iron in which the magnetic moments of the electrons spontaneously align to produce a net magnetization. Both ferromagnetism and superconductivity are phases of matter in which the tiny magnet moments, or 'spins' of the electrons (which can be either up or down), develop a highly correlated state. When two electrons have opposite spins they are said to be in a 'singlet' state. In conventional superconductivity the electrons form a condensate of Cooper pairs — a liquid of strongly overlapping pairs of electrons in singlet spin states. By contrast, when two electrons have the same spin alignment (such as both up or both down) they form a 'triplet' state. In a metallic ferromagnet the electrons are mobile, or 'itinerant', but their strict spin-alignment favours triplet over singlet pairing and conventional Cooper pairs are unable to form.

In ferromagnets, the electrons lower their interaction energy (and therefore become more stable) by taking advantage of the Pauli exclusion principle, which prevents elec-

Renato Paro is at the Center for Molecular Biology (ZMBH), University of Heidelberg, Im Neuenheimer Feld 282, 69120 Heidelberg, Germany. e-mail: paro@sun0.urz.uni-heidelberg.de

1. Turner, B. M. *Bioessays* (in the press).
2. Strahl, B. D. & Allis, C. D. *Nature* **403**, 41–45 (2000).
3. Rea, S. *et al.* *Nature* **406**, 593–599 (2000).
4. Jenuwein, T. *et al.* *Cell. Mol. Life Sci.* **54**, 80–93 (1998).
5. Hsu, J. Y. *et al.* *Cell* **102**, 279–291 (2000).
6. Lo, W.-S. *et al.* *Mol. Cell* **5**, 917–926 (2000).
7. Cheung, P. *et al.* *Mol. Cell* **5**, 905–915 (2000).

trons with parallel spins from coming too close to each other. Conventional superconductivity is driven by quite a different effect: here, electrons passing through the lattice of the crystal perturb the atoms around them. These lattice vibrations cause the electrons to attract each other to form Cooper pairs. In this way, conventional superconductivity is mediated by the tiny lattice vibrations or 'phonons' that accompany the motion of the electrons.

In the 1960s, a new kind of electron pairing mechanism was proposed that relied on magnetic interactions rather than lattice vibrations. In this unconventional superconducting state the particles are spin aligned to form triplet pairs. Such a paired state was originally suggested to explain the superfluidity of liquid helium-3 (refs 2–4). As in a ferromagnet, the formation of spin-aligned Cooper pairs lowers the repulsive energy of the medium by keeping the particles further apart. It seemed reasonable to suggest that the same forces that favour spin alignment in a ferromagnet might also give rise to spin-triplet pairing in 'nearly ferromagnetic' metals^{5,6} — that is, metals on the verge of undergoing a transition to a ferromagnetic state at low temperatures. Unlike conventional superconductivity, the glue that binds electrons into triplet Cooper pairs is not phonons, but magnetic spin fluctuations.

Such 'magnetically mediated' pairing requires strong spin fluctuations, such as those that develop in materials on the brink of magnetic order. Although magnetically mediated superconductivity was originally proposed for the 'nearly ferromagnetic' metal palladium, it was never observed. It was later discovered that spin-fluctuation pairing could also occur in a related class of materials on the border of antiferromagnetism, in which spins develop an alternating alignment. In such materials, electrons can lower their energy by forming even more complex Cooper pairs, such as the 'd-wave pairing' now known to exist inside high-

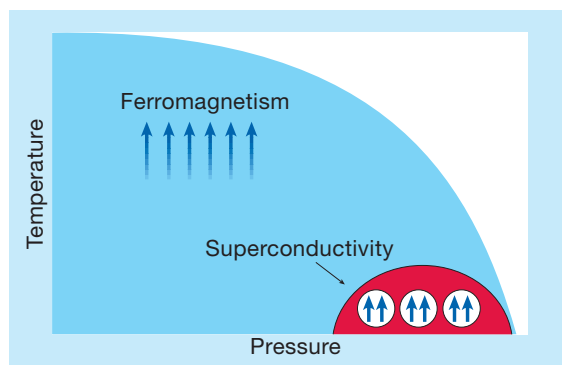


Figure 1 Temperature–pressure phase diagram for UGe_2 . Superconductivity is observed in a narrow regime just within the ferromagnetic state of an itinerant electron system. Arrows indicate relative spin orientations. (Courtesy of K. Ahilan, S. S. Saxena & G. G. Lonzarich.)

temperature superconductors. But despite 30 years of speculation, there were still no firm examples of superconductivity having developed in association with ferromagnetic spin fluctuations.

The discovery by Saxena *et al.*¹ of superconductivity inside a ferromagnet (a uranium–germanium compound, UGe_2) finally sets the record straight. Saxena *et al.* realized that past experiments simply did not get close enough to the boundary with magnetism. Unless the spin-pairing forces are strong, tiny amounts of disorder are enough to destroy the coherent scattering effects needed for triplet pairing. The idea was to use high pressures to bring a ferromagnet to the very brink of magnetic order. The temperature at which a ferromagnet loses its ferromagnetism is the Curie point, and applying pressure to a ferromagnet depresses its Curie temperature to zero kelvin. Saxena *et al.* reasoned that when the Curie temperature goes to zero, the resulting strong magnetic fluctuations would eventually cause the system to develop the long-sought-after superconductivity.

This idea had previously been tried without success on another ferromagnet, manganese silicon. Here it turns out that a quirk of crystal symmetry — the absence of inversion symmetry — prevents triplet pairs from forming. The material chosen for this study, UGe_2 , avoids this problem, and can also be prepared in crystals of high purity. The *f*-electron character of this ferromagnet causes its Curie temperature to be highly sensitive to external pressure, and 1.6 gigapascals (1.6 GPa or 16 kilobars), well within the range of modern high-pressure experiments, is sufficient to eliminate the magnetism.

But Saxena *et al.* report another surprise: they find that superconductivity actually develops at a lower pressure, before the ferromagnetism is eliminated, producing an unexpected situation in which superconductivity coexists with ferromagnetism. Measurements of the a.c. susceptibility confirm the observation of bulk superconductivity within the ferromagnet. Could the observed superconductivity be of the old-fashioned antiparallel, singlet variety? It's highly unlikely. The authors point out that the energy bands of spin-up and spin-

down electrons inside the ferromagnet are likely to be split by an energy gap of roughly 70 millielectronvolts — equivalent to a temperature of almost 1,000 kelvin. The only way in which Cooper pairing could take place under these conditions is for the spins in the Cooper pair to be aligned in a parallel, or triplet, configuration⁶.

What then is the origin of the electron pairing in UGe_2 ? No definitive answer can yet be given, but the discovery of pairing in the region where magnetism vanishes is a strong clue that it is likely to be magnetic. Furthermore, Saxena *et al.* show that the highest superconducting-transition temperature is obtained at a pressure where the density of electron states is at its highest, and it is here that magnetically mediated pairing is expected to be strongest. Unfortunately, the ferromagnetic phase ends abruptly at 1.6 GPa, and no trace of the superconductivity survives at higher pressures. Future experiments that combine pressure with an applied magnetic field may show how superconductivity extends into the non-magnetic phase. Moreover, experiments using the much higher pressures available with diamond anvil pressure cells may be able to induce superconductivity on the verge of magnetism in a broad variety of magnetic materials. ■

Piers Coleman is in the Department of Physics and Astronomy, Rutgers University, 136 Frelinghuysen Road, Piscataway, New Jersey 08854-8019, USA. e-mail: coleman@physics.rutgers.edu

1. Saxena, S. S. *et al.* *Nature* **406**, 587–592 (2000).
2. Brueckner, K. A., Soda, T., Anderson, P. W. & Morel, P. *Phys. Rev.* **118**, 1442–1446 (1960).
3. Nakajima, S. *Prog. Theor. Phys.* **50**, 1101–1109 (1973).
4. Brinkman, W. F., Serene, J. W. & Anderson, P. W. *Phys. Rev. A* **10**, 2386–2394 (1974).
5. Berk, N. F. & Schrieffer, J. R. *Phys. Rev. Lett.* **17**, 433–435 (1966).
6. Fey, D. & Appel, J. *Phys. Rev. B* **22**, 3173–3182 (1980).

Erratum

The technique known as protein signature analysis, developments of which were discussed by Satish K. Nair and Stephen K. Burley in the article "Functional genomics: Recognizing DNA in the library" (*Nature* **404**, 715–718; 2000), was originally described in a paper by T. W. Muir, P. E. Dawson, M. C. Fitzgerald and S. B. Kent in *Chemistry and Biology* (**3**, 817–825; 1996).

Daedalus

Spinning and leaking

Can microwaves damage living tissue — in particular the brain tissue of mobile-phone users? Mobile phones put out too little power to cause much heating; even so, they may induce more subtle non-thermal damage. Daedalus now proposes a mechanism.

A typical cell membrane is a fairly fluid bilipid layer with a few big protein molecules embedded in it. These molecules are receptors or 'antennae' by which the cell responds to specific molecules around it, or announces its internal state to the outside world. Every protein molecule, says Daedalus, is polar, and will tend to align itself with an electric field. It is also chiral, and often helical: rotate it in its socket, and it will screw inwards or outwards. Together, these facts mean that a rotating electric field could screw it into or out of the cell.

Almost any microwave source, such as a mobile-phone transmitter, will have some component of circular polarization. It will make billions of rotations every second. Any volume of brain tissue must have a few cells bearing protein molecules that happen to resonate at its rotation frequency. Even very weak coupling to the field should soon spin these molecules through the few turns needed to screw them out of the cell membrane. The cell might then simply leak to death — penicillin kills bacteria by making them leak. Or it might plug the hole with some wrong protein, creating more indirect trouble. Either way, the cell could lose the information entrusted to it, causing the memory loss complained of by mobile-phone users.

So, says Daedalus, to reduce the hazards of mobile phones, minimize the component of circularly polarized radiation in their output. But he also sees the way to a powerful new radiotherapy. If you knew the rotational resonant frequency of a particular protein in a particular type of cell, you could hit it with circularly polarized microwaves of exactly that frequency. Molecules of that protein would be spun out of their cell membranes, killing the cells or inducing them to take up some specially tailored therapeutic 'channel-blocker' which you had cunningly injected beforehand. Other proteins, and other cells nearby, would be unaffected. This elegant therapy could be directed against foreign bacteria or parasites in the body, or over- or underperforming glands, or cells succumbing to viral or neoplastic derangement. It would need much knowledge and insight to develop, but would have amazing power, specificity and freedom from side effects.

David Jones

Obituary

W. David Kingery (1926–2000)

In these days when the term ‘paradigm shift’ is applied with apparent abandon to any new set of data, it is salutary to recall that genuine revolutions in science and its disciplinary structure can and do take place. And it is reassuring to know that these revolutions can still arise from the work of an individual. One such scientist was David Kingery, who died on 29 June at the age of 73.

The paths to revolution take many forms. One route is the building of a rich and varied but consistent body of work, coupled with the writing of a ‘founding treatise’ (textbook is too mild a word; the absence of a precise term bespeaks the rarity of the event) which sets that body of work within a totally new vision of the subject. The method applies beyond science: from Ruy Lopez on chess to Machiavelli on ‘modern’ politics, the founding treatise brings a permanent change in concept. Each subject has its examples (I have a Japanese colleague who is forever seeking first editions of Charles Coulson’s or Linus Pauling’s works). For ceramics, the treatise is Kingery’s comprehensive but modestly titled *Introduction to Ceramics*, first published in 1960 (the year in which the photograph here was taken).

In the 1950s, as a consequence of long traditions of technological and artistic development, ceramics were still seen as distinctive products of individual industries. Heavy clays (for building), steelplant refractories (for reaction containment at high temperature), glass, pottery and porcelain, each had its separate culture. Such were the uncertainties of manufacture that each also had its body of folklore. At the same time, progress in solid-state chemistry and physics (notably in the work at Philips, Eindhoven, on the ceramic magnets, the spinels) was showing the dramatic potential for the application of ceramic materials to sectors outside the traditional industries.

Kingery entered this arena in 1948, when he began his doctoral studies at the Massachusetts Institute of Technology (MIT) at the invitation of F. H. Norton. Over the following years, as he made the rite of passage through to full professor at MIT in 1962, he made prolific contributions to the literature of property measurement (notably thermal conductivity) and the processing of ceramics (notably sintering — the bonding together of powder particles during heat treatment to yield dense, shaped



Scientist who led a revolution in ceramics

products). At the same time he was producing books: *Property Measurement at High Temperatures*, *Kinetics of High Temperature Processes*, *Ceramic Fabrication Processes*, all leading to the *Introduction to Ceramics*.

The quantitative models that still bear Kingery’s name, such as that for the most widespread method of consolidating shaped assemblies of powders, illustrate the factors that have made his work so influential. The new slant on the problem, the systematic reduction to the essential elements, the elegance of the analysis, the precise matching of theory to verifiable parameters — these were all qualities brought to bear as a subject, which (in his own words) had been “akin to a craft industry”, was transformed to “an industry based on engineering science”.

Kingery’s guiding theme was that ceramics should be seen as a class of materials with common and controllable attributes rather than as the products of isolated and idiosyncratic cultures. He advanced this theme with a compelling mixture of conviction, logic and evidence (largely from his own research). The resulting change in the subject has been radical and permanent, pervading all subsequent developments from ceramic engine components to high-temperature superconductors.

The extent of the revolution was recognized by Kingery’s receipt of an array

of prizes and honours, culminating in 1999 with Japan’s highest private award, the Kyoto Prize. Japan has shown great innovative flair in developing ceramic products, from devices used in electronics and communications, through to domestic articles. So the decision to credit Kingery for laying the required scientific foundations was informed by the most direct of experience.

The ready acceptance of his research contributions owed much to his personal style. Three aspects deserve emphasis. The first was the authority established by a remarkable capacity for work. The second was the determination to operate within a logical framework which itself drew strength from the wider context. At first these links were to the neighbouring disciplines of solid-state science and of engineering. Later, they broadened to cover the cultural context of technological development, extending to the archaeological record and most recently to the integration of technology in social development (his move in 1988 to the University of Arizona was to a joint appointment in materials science and anthropology).

The third aspect was a splendid, if highly personal, style of communication. Kingery’s syntax could be refined, even opulent — his ability in speaking to bring sentences of remarkable complexity to an elegant conclusion was a challenge to all who shared platforms with him. Yet the meaning always shone through.

A further achievement was Kingery’s development of a graduate teaching programme at MIT, the influence of which is felt to this day. The natural tendency of graduate students to mythologize their mentors was in his case intensified by the range and glamour of his life outside science: ocean sailing, flying, riding and the nurturing of a spin-off company were all carried out with unnerving charisma. As those students came to maturity, however, it was the generosity and friendship that were remembered.

I last met David Kingery in May, at the annual meeting of the American Ceramic Society which had been so much his home. Fleeing from the perils of conversation devoted to “bridge, golf or stocks”, he was intent on replacing “the naïve use of 1600–1950 physics as the paradigm defining science”. He will be sorely missed, as a friend, as a colleague, and as a wise and imaginative guide to the expansion of horizons.

Richard J. Brook

Richard J. Brook is at the Engineering and Physical Sciences Research Council, Polaris House, North Star Avenue, Swindon SN2 1ET, UK. e-mail: richard.brook@epsrc.ac.uk

Predicted vCJD mortality in Great Britain

Modelling the latest data puts a ceiling on the likely number of vCJD cases.

There is continued speculation about the likely number of cases of variant Creutzfeldt–Jakob disease (vCJD) that will occur in Great Britain in the wake of the BSE epidemic in cattle and in light of a recent cluster of vCJD cases in Leicestershire, England. We show here that the current mortality data are consistent with between 63 and 136,000 cases among the population known to have a susceptible genotype (about 40% of the total population), with on average less than two cases of vCJD arising from the consumption of one infected bovine.

Between 1980 and, roughly, 1996, about 750,000 cattle infected with BSE were slaughtered for human consumption in Great Britain^{1,2}. The recent fall in the annual incidence of vCJD (3, 10, 10, 18 and 14 deaths in 1995–99, respectively) has been interpreted as showing the disease to be in decline³. However, the suggestion by the European Union Scientific Steering Committee that up to 500,000 people could have been exposed to BSE from a single infected bovine⁴ has fuelled speculation that millions of consumers are at risk.

We used extensive scenario analyses to relate the number of infected cattle slaughtered for consumption^{1,2} to the incidence of vCJD stratified by time and age⁵, assuming that only those with the observed susceptible genotype (having two copies of the methionine codon at codon 129 of the *PrP* gene) are at risk (other genotypes may be resistant to infection or have longer average incubation periods). We explored a wide range of assumptions regarding the distribution of the vCJD incubation period, the relative infectivity of cattle by incubation stage, and the effectiveness of control measures at reducing human exposure to infected material⁵. The scenarios result from the exploration of over 5 million combinations of parameters.

From the vCJD incidence in 1999, we can substantially reduce the uncertainty in the upper bound of the total number of cases among people with the susceptible genotype. Our earlier analyses showed that if

Figure 1 The range of vCJD epidemic scenarios consistent with the time- and age-stratified vCJD mortality data to the end of 1999.

a, The age distribution (at death) of the 53 cases of vCJD confirmed to the end of 1999, as of 1 March 2000 (R. Will, personal communication). **b**, The relationship between mean incubation period (conditional on the incubation period being ≤ 91 years), the average number of vCJD cases arising from consumption of one late-stage (last 6 months of the incubation period) bovine and the total epidemic size. The probability that an individual develops clinical disease at time u and age a is $p(u,a) = S(u,a) \int_{u-a}^{\infty} f(u-t, a-u+t) / (t, a-u+t) \exp(-\int_0^t l(t', a-u+t) dt') dt$, with $S(u,a)$ being the survival probability, $f(u,a)$ the incubation period distribution and $l(t,a) = \nu(t) \beta g(a) \int \Omega(z) w(z,t) dz$ the infection hazard; $\nu(t)$ is the effectiveness of control measures, β the transmission coefficient, $g(a)$ age-dependent susceptibility/exposure, $\Omega(z)$ the relative infectiousness of bovines time z before disease onset, and $w(z,t)$ the proportion of cattle slaughtered at time t and time z away from disease onset (estimates taken from ref. 2). We assume that the infectiousness of bovine tissues increases exponentially, doubling in the last 6 months of the bovine incubation period. We explored two forms for the age-dependent susceptibility/exposure — a uniform distribution and a normal distribution truncated at zero. We modelled the age-dependent incubation period using the scaling $f(u,a) = h(u)/s(a)$, where $h(u)$ is an age-independent incubation period and $s(a) = (\alpha_2 \exp(-\alpha_3 a) + \alpha_1) / (\exp(-\alpha_3 a) + \alpha_1)$. Each point represents a model scenario that is consistent at the 95% level with the marginal time- and age-stratified mortality data to the end of 1999 as of 1 March 2000.

fewer than 15 cases arose in 1999 in any age group, the upper bound on the number of cases in this population was approximately 500,000 (ref. 6). This is reduced to about 136,000 when the incidence in 1999 is combined with the observed stable age structure (Fig. 1a). Our model suggests that the age distribution of vCJD cases could not have arisen from an age-dependent incubation period alone (which would give rise to an increasing median age over time), and thus

the under-40s must be either more susceptible to infection or have been more highly exposed to the aetiological agent.

The current data suggest that, on average, no more than two cases of vCJD could arise from the consumption of one maximally infectious bovine (Fig. 1b) — a year ago, this value's upper bound was over 100. This suggests a substantial species barrier, given that thousands of people might eat material from a single animal⁴. Furthermore, a large number (more than 6,000) of cases can only arise if the mean incubation period is about the same length as average life expectancy (Fig. 1b; Table 1). If the average annual incidence of vCJD over the next three years is fewer than 15 cases, then the maximum total number of cases would fall to approximately 20,000 (Table 1).

Recently, preliminary results from tests of tonsil and appendix material for the presence of the abnormal prion protein have demonstrated a null result (zero infections detected in 3,170 tissues)⁷. Interpretation of

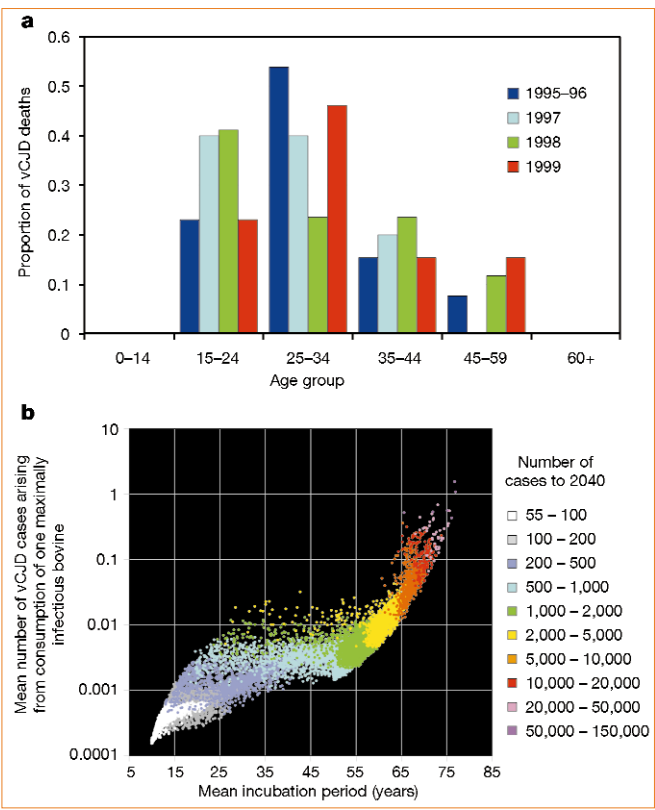


Table 1 Projections for the eventual total number of vCJD cases							
Mean i.p. (years)	Number of cases in 2000			Average number of cases in 2000–02			
	10–14	15–19	20 +	<10	10–14	15–19	20 +
<20	70–630	75–630	80–630	63–350	90–530	105–630	120–630
20–30	70–	87–	110–	70–	90–	105–	120–
	2,900	2,900	2,800	900	1,500	2,900	2,900
30–60	150–	150–	150–	150–	150–	150–	200–
	5,500	6,000	6,000	1,500	4,500	5,500	6,000
≥60	1,300–	1,300–	1,300–	1,300–	1,300–	1,300–	1,300–
	136,000	136,000	136,000	1,700	20,000	20,000	136,000

The range of total number of cases is given that is consistent with the number of cases recorded in 2000 and with the average annual number of cases in the next three years, calculated by the conditional mean incubation period (i.p.). Results include concentrated sampling of parameter combinations that result in large epidemics.

these results is complicated by a lack of knowledge of when in the incubation period infection becomes detectable. If tests can detect infection in the last 75% of the incubation period (with 100% sensitivity and specificity), then the upper bound on total epidemic size in the susceptible genotype population is reduced to 80,000. However, if tests can detect infection only in the last 50% of the incubation period, the current uncertainty remains. Future vCJD case data, combined with the ongoing retrospective and prospective studies of tonsil and appendix material, will continue to reduce the remaining uncertainty in the future course of vCJD incidence in the coming years.

Encephalopathies

Scrapie in Britain during the BSE years

The experimental transmission of bovine spongiform encephalopathy (BSE) to sheep¹ raised the possibility that some sheep in the United Kingdom could have been infected during the 1980s after exposure to BSE-contaminated feed. In contrast to new diseases that have appeared in a number of feline species and wild ungulates², the symptoms of BSE in sheep are very similar to another transmissible spongiform encephalopathy called scrapie, which has been endemic in Britain for over 200 years. Although so far no cases of BSE in sheep have been found, these may have been misdiagnosed as scrapie. Here we present data describing the historical changes in

Azra C. Ghani, Neil M. Ferguson, Christl A. Donnelly, Roy M. Anderson
Wellcome Trust Centre for the Epidemiology of Infectious Disease, University of Oxford, Oxford OX1 3FY, UK
e-mail: neil.ferguson@ceid.ox.ac.uk

1. Anderson, R. M. *et al.* *Nature* **382**, 779–788 (1996).
2. Ferguson, N. M., Donnelly, C. A., Woolhouse, M. E. J. & Anderson, R. M. *Phil. Trans. R. Soc. Lond. B* **352**, 803–838 (1997).
3. Newby, M. J. & Thomas, P. J. *Witness Statement no. 313B to the BSE Enquiry* (25 January 2000).
4. *Opinion of the Scientific Steering Committee on the Human Exposure Risk (HER) via Food with respect to BSE, European Union Scientific Steering Committee* (10 December 1999).
5. Ghani, A. C., Ferguson, N. M., Donnelly, C. A., Hagenaars, T. J. & Anderson, R. M. *Proc. R. Soc. Lond. B* **265**, 2443–2452 (1998).
6. Ghani, A. C. *et al.* *Proc. R. Soc. Lond. B* **267**, 23–29 (2000).
7. Ironside, J. W. *et al.* *Lancet* **355**, 1693–1694 (2000).

scrapie incidence³, and find no evidence for a peak in scrapie incidence before, during or after the BSE outbreak, making it unlikely that a substantial epidemic of BSE has occurred in the sheep population.

A postal survey of 11,554 sheep farms⁴ requested information on the length of time farmers had been involved in farming sheep and the year in which scrapie had first appeared in their flock. The response rate was 61%, and 15% of the relevant responders reported ever having had scrapie. For each year we calculated how many responders were farming, how many had not yet had any scrapie, and how many acquired their first case that year. If the force of infection per farm (FI) is the rate at which farms acquire cases, we define the first-case force of infection (FI₁) as the number acquiring a first case in a given year divided by the number of farms at risk of doing so.

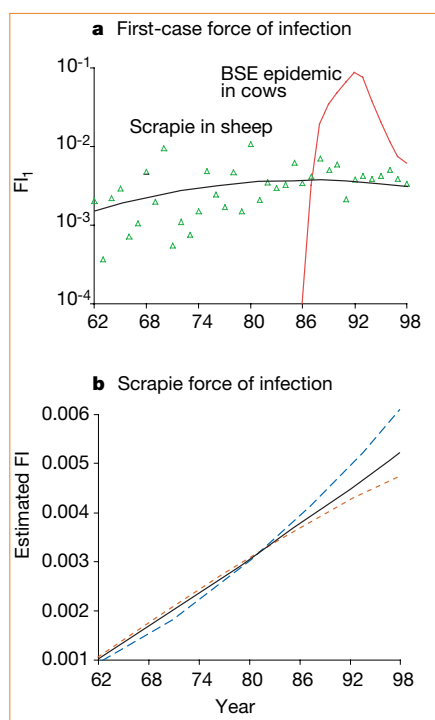


Figure 1 Analysis of results from scrapie survey among British farmers. **a**, Proportion of farms at risk of having a first scrapie case per year (FI₁). The fitted quadratic curve (estimated using logistic regression) was not sensitive to smoothing of the high proportion of cases reported in 1970 and 1980 that probably resulted from digit preference. Also shown is the proportion of cattle herds with no previous BSE that experienced a first case in each year. **b**, Estimation of scrapie FI for all farms (not just first case). If the probability of acquiring any case in year n is $\theta_n = \theta(1 + \alpha n + \beta n^2)$, the probability ϕ_n of a first case in year n , given no previous cases, can be expressed as

$$\phi_n = 1 - \left(\frac{c + m \left(n + \frac{\alpha n(n+1)}{2} + \frac{\beta n(n+1)(2n+1)}{6} \right)}{c + m \left((n-1) + \frac{\alpha n(n-1)}{2} + \frac{\beta n(n-1)(2n-1)}{6} \right)} \right)^{-c}$$

A gamma distribution for the probability density function of θ is assumed (with coefficient of variation = $\sqrt{1/c}$ approximated by 1.3, the current coefficient of variation in flock size). The parameters m , α (linear coefficient) and β (quadratic coefficient), and hence the force of infection θ per year, can be estimated from the reported proportion of first cases each year (ϕ_n). Parameter estimates (solid line) were $m = 0.0031$ (0.0022, 0.0039), $\alpha = 0.039$ (0.015, 0.062), $\beta = 6.6 \times 10^{-5}$ (–0.0018, 0.0019), showing no significant departure from linearity. Dotted lines give the estimated FI assuming the coefficient of variation in θ is $\pm 50\%$ the coefficient of variation for flock size.

Figure 1a shows FI₁ for 1962–98, together with the FI₁ for BSE in cattle herds for comparison. If an increase in spongiform encephalopathy in sheep was linked to the BSE epidemic, we might expect a substantial increase in new scrapie-affected farms during this time (or preceding the BSE epidemic by a few years owing to the shorter incubation period of BSE in sheep relative to cattle⁵). Instead, we see a steady increase from an estimated mean of 0.2% in 1970 to 0.4% in 1990, followed by a slight decline. Although a curve best fits the data, there is no sharp peak. This is apparent from the raw data. Further, the formal confidence interval for the date of maximum FI₁ is very wide.

If some farms are more susceptible to scrapie than others, analysis of the farms with first cases only (FI₁) will give a biased estimate of the true force of infection for all farms (FI). Variability in the probability of acquiring scrapie arises through a number of factors, including flock size and type⁶. We estimated the trend in FI from the survey data by expressing the probability of acquiring the first case in terms of the probability of acquiring any case and the variation in this probability between farms. FI was estimated to increase linearly over time (Fig. 1b). Factors that may underlie this trend include recall biases and increases in average flock size and scrapie awareness. Note that variation in farm susceptibility introduces a curved shape to FI₁ (as in Fig. 1a) even when the underlying probability of infection is linear. The recent apparent decline in first-case farms may therefore be misleading.

Overall, the data indicate that there is a gradual linear increase in the probability of farms acquiring a scrapie case over the past 36 years. Sensitivity analysis suggests that departures from this pattern would have been detected had the probability of having scrapie since 1981 increased to twice that expected from 1962–80. Different patterns in FI in a small subsample of farms may therefore be overlooked. Splitting the data by farm size does not reveal any such heterogeneity. A further subgroup to analyse would be farms that fed large amounts of meat and bone meal, but these farms were not identified in the survey.

The link between BSE and variant Creutzfeldt–Jakob disease in humans^{7,8} raises the question of whether and to what extent BSE entered the British sheep population. Mass screening⁹ to check for BSE has been proposed but is not yet feasible. The survey provides additional information that farms raising both sheep and cattle were not at a higher risk of having had scrapie, and regional scrapie and BSE incidences were not significantly correlated. Reports of case numbers in the past 12 months and previous 5 years were also consistent with a recent constant incidence of scrapie¹⁰. The data presented here cannot be used to

determine whether any sheep became infected with BSE, but they do suggest that the change in sheep spongiform encephalopathy incidence has been a gradual process, with no large epidemic during the 1980s and 1990s.

Mike B. Gravenor*, D. R. Cox †,
Linda J. Hoinville‡, Alies Hoek‡,
Angela R. McLean*

*Institute for Animal Health, Compton, Berkshire
RG20 7NN, UK

e-mail: michael.gravenor@bbsrc.ac.uk

†Department of Statistics, University of Oxford,
1 South Parks Road, Oxford OX1 3TG, UK

‡Epidemiology Department, Veterinary
Laboratories Agency, New Haw, Addlestone,
Surrey KT15 3NB, UK

1. Foster, J. D., Hope, J. & Fraser, H. *Vet. Rec.* **133**, 339–341 (1993).
2. Kirkwood, J. K. & Cunningham, A. A. *Vet. Rec.* **135**, 296–303 (1994).
3. Butler, D. *Nature* **395**, 6–7 (1998).
4. Hoinville, L., McLean, A. R., Hoek, A., Gravenor, M. B. & Wilesmith, J. *Vet. Rec.* **145**, 405–406 (1999).
5. Goldmann, W., Hunter, N., Smith, G., Foster, J. & Hope, J. *J. Gen. Virol.* **75**, 989–985 (1994).
6. McLean, A. R., Hoek, A., Hoinville, L. J. & Gravenor, M. B. *Proc. R. Soc. Lond. B* **266**, 2531–2538 (1999).
7. Collinge, J., Sidle, K. C., Meads, J., Ironside, J. & Hill, A. F. *Nature* **383**, 685–690 (1996).
8. Bruce, M. E. *et al.* *Nature* **389**, 498–501 (1997).
9. Hill, A. F. *et al.* *Neurosci. Lett.* **255**, 159–162 (1998).
10. Gravenor, M. B., Cox, D. R., Hoinville, L. J., Hoek, A. & McLean, A. R. *Proc. R. Soc. Lond. B* (submitted).

Physiology

Exercise and reduced muscle mass in starlings

Muscles are often viewed as force-producing structures that increase or decrease in size according to their activity^{1,2}. But an increased muscle mass, although desirable for extra power, may also impose unwanted costs. Here we show that flight-muscle mass in starlings induced to perform more take-off flights actually decreases as a result of exercise. Our findings indicate that birds can strategically regulate a lower muscle mass to make themselves lighter and so cut flying costs without compromising their flight performance. This suggests that muscle size may be influenced by factors other than workload.

The muscle use–disuse hypertrophy–atrophy hypothesis is often invoked to explain flight-muscle size in wild birds: for example, enlarged pectoralis muscle (the main force generator for powering flight) is seen as an indicator of increased flight demands^{3,4}. The costs of bearing a larger muscle mass have not been investigated, however. One such cost could be carrying an increased load during flight. We manipulated take-off flight exercise in European starlings, *Sturnus vulgaris*, and measured changes in musculature in response to flight.

Daily flight exercise led to a reduction in

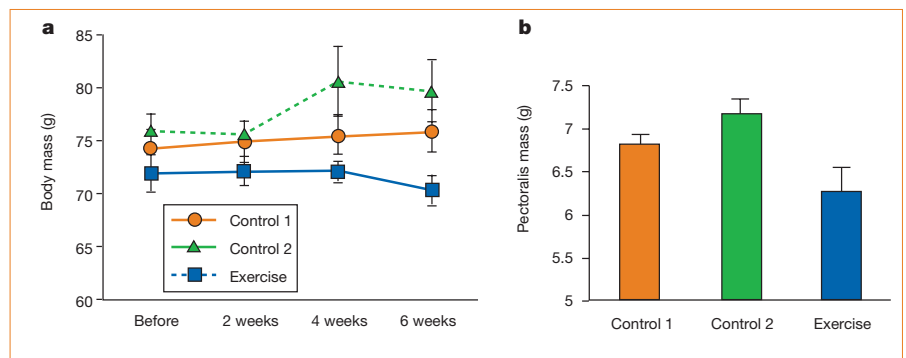


Figure 1 Morphological responses to exercise. **a**, Mean ($\pm$ s.e.) change in body mass during the experiment ($F_{6,45} = 2.35$, $P = 0.047$); and **b**, mean ($\pm$ s.e.) pectoralis muscle mass at the end of the experiment ($F_{2,15} = 3.95$, $P = 0.042$). Eighteen adult starlings were housed under controlled conditions based on an 8-h day and randomly allocated to three groups. 'Exercise' group birds were trained to fly between perches at opposite ends of a flight cage ($0.6 \times 0.8 \times 2$ m) by positive reinforcement with food rewards. Birds had 1-hour exercise trials on most days, with 34 trials over 6 weeks. First control group, control 1: birds were restrained in a small cage and fed *ad libitum* for the 1-h duration of the exercise trials; second control group, control 2: birds were restrained in a small cage and received the same food rewards at the same times as the exercising birds. On four occasions (the day before the experiment and then every two weeks), we recorded take-off force production (by tethered strain gauges), wing area, body mass and fat stores¹¹. After the trials, we measured pectoralis wet mass and fibre length¹²; supracoracoideus wet mass, fibre length and pinnation angle; and gastrocnemius muscle attributes. There was no difference in activity among groups when they were outside their cages. Further details are available from the authors.

body mass (Fig. 1a) and wing loading ($F_{6,45} = 2.28$, $P = 0.052$) compared with controls. Most birds had sparse fat deposits and there was no difference in fat score among treatment groups (Friedman $S_2 = 2.29$, $P = 0.319$). However, birds in the exercise group had reduced pectoralis muscle mass (Fig. 1b; pectoralis fibre length also decreased: sternal, $F_{2,15} = 3.49$, $P = 0.057$; thoracic, $F_{2,15} = 2.68$, $P = 0.101$). There was no effect of exercise on other musculature measured ($F_{2,15} < 1.85$, $P > 0.19$). Although differences in body mass and pectoralis mass appear to exist between control groups at 4 and 6 weeks (Fig. 1a, b), these differences were not significant ($F_{1,11} < 2.17$, $P > 0.171$). The reduction in body mass and pectoralis muscles in these birds did not affect take-off flight performance, which we measured as peak take-off acceleration ($F_{6,45} = 0.42$, $P = 0.859$).

Our results contradict the use–hypertrophy hypothesis. The higher nutritional requirements⁴ to fuel exercise cannot explain the changes in pectoralis mass, because daily access to food was not limited. Also, subcutaneous fat stores were not affected and the pectoralis comprises less than 1% fat^{2,5,6}. As sternal pectoralis fibre length decreased by 5% in the exercise group of birds, water loss cannot explain the decrease in muscle size.

Birds in the exercise group performed more flights than controls, so any strategy that reduces energetic flight costs must be adaptive. We propose that birds are regulating a lower body mass level to reduce flight costs, even at the expense of muscle size⁷. Starlings reduce their body mass in response to a reduced wing area⁸. The costs of decreasing pectoralis size are balanced by the benefits of losing 10% of body mass (M).

This idea is also interesting from a scal-

ing point of view: for a given wing area, decreases in pectoral/body mass and wing loading will result in a smaller decrease in force-generating ability (assuming a muscle force α to fibre area $\alpha M^{2/3}$). So a 10% decrease in pectoral/body mass would predict a 4.7% decrease in force generation.

We know that atrophied pectoral muscles retain a high density of mitochondria even though contractile proteins and fat stores are depleted², and that the pectoralis can undergo enzymatic adaptation to increase oxidative metabolism⁵. The smaller pectoralis may therefore be able to maintain its oxidative capacity. The relative abundance of the two fast-twitch muscle fibre types (dependent on either aerobic or anaerobic glycolysis) is unlikely to alter as a result of exercise because the starling pectoralis consists exclusively of fibres of the aerobic type⁹. These observations indicate that pectoralis size cannot simply be linked to functional performance.

The increased flight costs of losing pectoralis mass are offset by the decreased flight costs of losing overall body mass, resulting in a new equilibrium of body condition based on changes in flight demands. Therefore, muscle size is not merely a function of its activity or storage capacity, but also reflects costs of resource allocation within the whole body¹⁰. The size of flight muscle thus represents a dynamic cost–benefit trade-off of a broad range of potential limiting factors.

John P. Swaddle*, Andrew A. Biewener†

*Centre for Behavioural Biology, School of
Biological Sciences, University of Bristol,
Woodland, Bristol BS8 1UG, UK

†Concord Field Station, MCZ, Harvard University,
Old Causeway Road, Bedford, Massachusetts 01730,
USA

e-mail: abiewener@oeb.harvard.edu

- Holloszy, J. O. & Coyle, E. F. *J. Appl. Physiol.* **56**, 831–838 (1984).
- Gaunt, A. S., Hikida, R. S., Jehl, J. & Fembert, L. *Auk* **107**, 649–659 (1990).
- Marsh, R. L. *Physiol. Zool.* **57**, 105–117 (1984).
- Austin, J. E. & Fredrickson, L. H. *Auk* **104**, 694–699 (1987).
- Driedzic, W. R., Crowe, H. L., Hicklin, P. W. & Septon, D. H. *Can. J. Zool.* **71**, 1602–1608 (1993).
- O'Connor, T. P. *J. Comp. Physiol. B* **165**, 298–305 (1995).
- Brown, R. E. & Saunders, D. K. *Can. J. Zool.* **76**, 26–32 (1998).
- Swaddle, J. P. & Witter, M. S. *Can. J. Zool.* **75**, 1135–1146 (1997).
- Rosser, B. W. C. & George, J. C. *Can. J. Zool.* **64**, 1174–1185 (1986).
- Witter, M. S. & Cuthill, I. C. *Phil. Trans. R. Soc. Lond. B* **340**, 73–90 (1993).
- Helms, C. W. & Drury, W. H. *Bird Band* **31**, 1–40 (1960).
- Biewener, A. A., Dial, K. P. & Goslow, G. E. *J. Exp. Biol.* **164**, 1–18 (1992).

Materials

Peeling and sharpening multiwall nanotubes

To realize the full potential of multiwall carbon nanotubes in applications such as biological and scanned probes, it is desirable to develop techniques for controlling their shape and geometry. Here we describe a method by which the outer layers of a multiwall nanotube can be successively removed at the end to produce what is effectively a sharpened structure.

Carbon nanotubes, owing to their unique mechanical and electrical properties, are candidates for a host of applications such as catalysts¹, biological cell electrodes², nanoscale electronic³ and mechanical⁴ systems, and scanned probe microscope and electron field emission tips^{5,6}. Many of these applications would be facilitated by some tailoring of the nanotube. For example, an 'ideal' scanned probe, field emission or biological electrode tip should be long, stiff and tapered for optimal mechanical response and have an electrically conducting tip. In addition, it would be useful to be able selectively to expose nested concentric nanotubes in a nanobearing. Techniques exist for growing nanotubes at preselected sites⁷ and for modifying nanotube ends through chemical etching⁸, but not for finely controlled shaping of the tubes.

We have found a simple and reliable method that allows controlled engineering or shaping of multiwall carbon nanotubes and enables average multiwall nanotubes to be easily converted into tips with ideal geometry for scanned probe, field emission, biological insertion or mechanical nanobearing applications. The shaping process involves the electrically driven vaporization of successive layers (that is, the tube walls) of the multiwall nanotube, with outer layers being removed in turn near the end of the nanotube, leaving the core nanotube walls intact and protruding from the bulk of the nanotube. This peeling and sharpening process can be applied repeatedly to the

same multiwall nanotube until the innermost tube(s) of the smallest diameter protrude, often with a tip having a radius of curvature comparable to that of one single-walled nanotube.

We demonstrate the method in a transmission electron microscope (TEM) configured with a custom-built mechanical/piezo manipulation stage with electrical feed-throughs to the sample. Figure 1 shows high-resolution TEM images of a conventional arc-grown multiwall carbon nanotube at different stages in the peeling and sharpening process. The left end of the nanotube (not seen in the image) is attached to a stationary zero-potential gold electrode. To the right (also not shown) is a larger nanotube that serves as the 'shaping' electrode: it is attached to the manipulator, whose potential can be controlled externally.

Figure 1a shows the nanotube in its pristine, as-grown state. In Fig. 1b, the shaping electrode has been momentarily brought into contact with the nanotube and a carbon onion has been inadvertently transferred from the shaping electrode to the nanotube, but the applied voltage (2.4 V) and current (170 μ A) are below the shaping threshold and no peeling or sharpening has taken place. Figure 1c shows what happens when the shaping electrode is brought into intimate contact with the tip of the nanotube at 2.9 V and 200 μ A: almost immediately, many layers of the nanotube are peeled away near its end and it now has a stepped diameter and is significantly sharpened. The carbon onion has been displaced to a benign position further down the tube. The newly exposed tip of the nanotube appears to be undamaged.

For Fig. 1d, the peeling and sharpening process has been repeated, resulting in a multiwall nanotube with highly desirable characteristics for many practical applications. The dominant protruding segment now consists of a three-walled electrically conducting nanotube with a radius of just 2.5 nm. Although we used an *in situ* TEM configuration to follow the sharpening and peeling process, this is not essential as the process could be performed blind and monitored only from the electrical characteristics of the nanotube. In addition, the 'shaping' electrode can readily be replaced by any conventional conducting substrate.

The physics behind our novel peeling and shaping process is intriguing. It is unlikely that uniform Joule heating of the nanotube would result in the observed behaviour. It is more likely that multiwall nanotubes conduct ballistically⁹, and the energy to break the carbon bonds and remove the nanotube layers originates from highly localized dissipation at defect scattering sites, located primarily at the ends of the tube. The ensuing avalanche of dissipation and bond-breaking would then lead to cata-

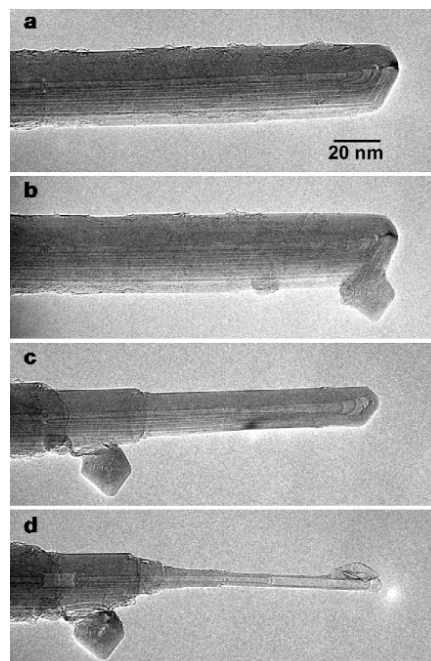


Figure 1 Transmission electron microscope images of a multiwall carbon nanotube being shaped. **a**, Nanotube in its pristine form: it contains approximately 37 walls and has an outer radius of 12.6 nm. **b**, A carbon onion has been inadvertently transferred to the nanotube end from the shaping electrode, but no attempt has been made to shape the nanotube. **c,d**, Results of the subsequent peeling and sharpening processes: the onion has simultaneously been displaced to a benign position down the tube axis. The shaped, or 'engineered', nanotube in **d** is thick and mechanically rigid along most of its length (not seen in the image), but tapers stepwise to a fine sharp tip that is electrically conducting and ideal for scanned probe microscopy or electron field emission applications. The final long nanotube segment contains three walls and has an outer radius of 2.1 nm.

strophic failure over a significant portion of the nanotube shell.

The fact that only the outer layers are affected indicates that electrical current in multiwall nanotubes may flow mainly in the outer carbon layers of the tube, in agreement with conclusions from magnetotransport experiments¹⁰.

John Cumings, Philip G. Collins, A. Zettl

Department of Physics, University of California at Berkeley, and Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA
email: azettl@physics.berkeley.edu

- Dillon, A. C. *et al. Nature* **386**, 377–379 (1997).
- Britto, P. J., Santhanam, K. S. V. & Ajayan, P. M. *Bioelectrochem. Bioenerg.* **41**, 121–125 (1996).
- Collins, P. G., Zettl, A., Bando, H., Thess, A. & Smalley, R. E. *Science* **278**, 100–103 (1997).
- Iijima, S. in *Proc. IEEE Eleventh Annual International Workshop on Micro Electro Mechanical Systems* 520–525 (IEEE, Heidelberg, Germany, 1998).
- Dai, H., Hafner, J. H., Rinzler, A. G., Colbert, D. T. & Smalley, R. E. *Nature* **384**, 147–150 (1996).
- Rinzler, A. G. *et al. Science* **269**, 1550–1553 (1995).
- Kong, J. *et al. Nature* **395**, 878–881 (1998).
- Tsang, S. C., Chen, Y. K., Harris, P. J. F. & Green, M. L. H. *Nature* **372**, 159–162 (1994).
- Frank, S., Poncharal, P., Wang, Z. L. & De Heer, W. A. *Science* **280**, 1744–1746 (1998).
- Bachtold, A. *et al. Nature* **397**, 673–675 (1999).

Superconductivity on the border of itinerant-electron ferromagnetism in UGe_2

S. S. Saxena^{*†‡}, P. Agarwal^{*}, K. Ahilan^{*}, F. M. Grosche^{*‡}, R. K. W. Haselwimmer^{*}, M. J. Steiner^{*}, E. Pugh^{*}, I. R. Walker^{*}, S. R. Julian^{*}, P. Monthoux^{*}, G. G. Lonzarich^{*}, A. Huxley[§], I. Sheikin[§], D. Braithwaite[§] & J. Flouquet[§]

^{*} Department of Physics, Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK

[†] Materials Science Centre, University of Groningen, Nijenborgh 4, 9747AG, The Netherlands

[§] Département de Recherche Fondamentale sur la Matière condensée - SPSMS, CEA Grenoble, 17 Av. des Martyrs, Grenoble 38054, France

The absence of simple examples of superconductivity adjoining itinerant-electron ferromagnetism in the phase diagram has for many years cast doubt on the validity of conventional models of magnetically mediated superconductivity. On closer examination, however, very few systems have been studied in the extreme conditions of purity, proximity to the ferromagnetic state and very low temperatures required to test the theory definitively. Here we report the observation of superconductivity on the border of ferromagnetism in a pure system, UGe_2 , which is known to be qualitatively similar to the classic d-electron ferromagnets. The superconductivity that we observe below 1 K, in a limited pressure range on the border of ferromagnetism, seems to arise from the same electrons that produce band magnetism. In this case, superconductivity is most naturally understood in terms of magnetic as opposed to lattice interactions, and by a spin-triplet rather than the spin-singlet pairing normally associated with nearly antiferromagnetic metals.

The origin of the remarkable stability or rigidity of structures in nature is a question of universal interest. The stability of simple systems of particles, such as atoms and molecules, is now described in great detail in terms of elementary quantum mechanics. But in the more complex systems that are of interest in condensed-matter physics and beyond, our understanding of rigidity is often incomplete or lacking entirely. A particularly dramatic example of large-scale quantum rigidity is the phenomenon of superconductivity, which is thought to arise from the emergence of an attractive interaction that in some sense overwhelms the usual Coulomb repulsion between pairs of electrons. In the standard model due to Frohlich and to Bardeen, Cooper and Schrieffer (BCS), the crucially important attraction arises naturally from the indirect effects of the deformable underlying crystalline lattice of ions¹. This BCS picture, in which electrons form bound pairs as a result of lattice interactions, is now believed to account well for the great majority of known superconductors. But there is a growing number of metallic compounds, including the high-transition-temperature superconductors, in which superconductivity appears anomalous and where the precise mechanism of electron pairing remains controversial.

Soon after the advent of the BCS theory, alternative models for electron pairing were proposed that relied directly on subtle dynamical effects of the electrons themselves, in a static, non-deformable lattice. An effective attraction between pairs of electrons, or more precisely of fermion quasiparticles near the Fermi surface, arises—at first sight, paradoxically—from the cooperative effects of a collection of electrons that mutually repel each other as they move over the underlying static lattice of ions. In contrast to the bare Coulomb repulsion, which is independent of the electron spin, a part of the complex interaction between the quasiparticles can depend on the relative orientation of the spins and thus of the magnetic moments of the carriers^{2–17}. In the simplest case of nearly ferromagnetic metals—that is, metals on the verge of undergoing a

transition to a ferromagnetic state at low temperatures—pairs of quasiparticles with parallel spins can attract while pairs with antiparallel spin tend to repel. When magnetic interactions in this example dominate over other types of quasiparticle interactions, parallel-spin quasiparticles tend to form pairs that must necessarily be in odd-parity orbitals. This can lead to spin-triplet, magnetically mediated, superconductivity^{2–8}. The effective magnetic interaction that we are describing is not the familiar dipole–dipole interaction in the theory of electromagnetism, which is relativistic and usually weak, but is a consequence of the Coulomb interaction itself, together with the subtle effects of quantum correlations; hence this effective magnetic interaction can be relatively strong, and potentially important for superconductivity in some cases.

Current theory suggests that this type of superconductivity is most likely to occur in metals that satisfy at least the following three conditions. First, they should be close to the border of ferromagnetism, either in a strongly paramagnetic or a weakly ferromagnetic state at low temperature where the longitudinal magnetic susceptibility that enters the magnetic interaction potential is strong and where magnetic interactions overwhelm other competing interactions. The balance between competing interactions, and also between the pair-forming and pair-breaking tendencies of magnetic interactions themselves, is a delicate one; superconductivity may not arise at all in some cases, or may exist in practice only over a very narrow range in a control parameter, such as hydrostatic pressure, used to tune the system towards the border of ferromagnetism. Second, the specimens selected must be of sufficiently high purity that the carrier mean free paths (due to either spin-dependent or spin-independent scattering mechanisms) exceed the typical dimensions of the spin-triplet pair states, that is, the characteristic superconducting coherence length. In contrast to conventional superconductors, these states are anisotropic in space and thus can be very sensitive to impurity scattering that tends to be essentially isotropic and strongly pair-breaking. Third, most candidate materials available, even at optimal lattice density and in their ideally pure states, may have to be cooled to the millikelvin temperature range to exhibit a spin-triplet form of magnetically

[‡] Present address: MPI Chemische Physik fester Stoffe, Bayreuther Str. 40, 01189 Dresden, Germany (F.M.G.); Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK (S.S.S.).

mediated superconductivity. Numerical analyses suggest that the superconducting transition temperature T_{SC} due to magnetic interactions is normally much lower for the spin-triplet pairing appropriate for our case of a nearly ferromagnetic metal, than for the spin-singlet even-parity pairing expected, for example, for the more complex case of a nearly antiferromagnetic metal at least in two dimensions¹⁸. (This assumes that the parameters entering the model are otherwise the same.) A key difference between these two cases is that an important factor defining the magnetic interaction potential—namely, the expectation of the inner product of the spins of the interacting quasiparticles—is three times stronger in magnitude for the spin-singlet state than for the spin-triplet state. This peculiar quantum property holds only for particles, such as the fermions of interest here, with spin quantum number of one-half in normal isotropic space.

Although there is growing evidence for the existence of magnetically mediated superconductivity in general, most of the information gathered thus far concerns the more complex examples of spin-singlet pairing normally associated with metals on the border of antiferromagnetism as opposed to ferromagnetism^{9–32}. Magnetic pairing of quasiparticles similar to the simplest kind we consider above is indeed thought to be relevant to superfluidity in liquid ³He (refs 3–7), but the corresponding phenomena in the nearly ferromagnetic metals has been more elusive. The lack of success in many searches for this simplest kind of magnetically mediated superconductivity, starting with early work on the nearly ferromagnetic metal Pd, has cast doubt on the validity of the theory of magnetically mediated superconductivity as it is presently formulated; even the possibility of a purely electronic mechanism for superconductivity in general has been doubted.

On closer examination, however, we find that very few itinerant-electron ferromagnets studied to date have been prepared in a sufficiently pure state, or have been ‘tuned’ (for example, by hydrostatic pressure) to be sufficiently close to the border of ferromagnetism, or cooled to sufficiently low temperatures, to provide a definitive check of the predictions of theory (for a recent review, see ref. 33). Recent analyses suggest that Pd, long considered the archetypal incipient ferromagnet, is in fact too far removed from the border of ferromagnetism to be a prime candidate for magnetically mediated superconductivity (P.M. and G.G.L., unpublished results). Low-temperature ferromagnets such as ZrZn₂ have not yet been prepared in sufficiently pure form to be expected to exhibit this kind of superconductivity, even near the critical

pressure where the Curie temperature T_C vanishes³⁴. In one promising system, MnSi, in which the three conditions mentioned above have apparently been met, the anticipated odd-parity spin-triplet superconductivity is not in fact observed^{35,36}. However, this material may fail to meet another and more subtle criterion needed for spin-triplet pairing. In particular, its B20 crystal structure lacks the inversion symmetry required to guarantee that equal-spin states of opposite momentum are degenerate and hence effectively coupled by magnetic interactions. In the nearly magnetic metal Sr₂RuO₄, evidence for spin-triplet superconductivity is mounting, but its connection with ferromagnetic spin susceptibility and the even stronger antiferromagnetic spin susceptibility inferred from neutron-scattering experiments has not yet been clarified^{37–39}. In RuSr₂GdCu₂O₈ superconductivity and magnetism with a ferromagnetic component appear to coexist at least in different layers of the layered perovskite structure^{40,41}. However, superconductivity is thought to arise from spin-singlet *d*-wave pairing in the nearly antiferromagnetic CuO₂ layers, and not from spin-triplet pairing associated with the ferromagnetism of the RuO₂ layers. Finally, in the system of nominal composition Y₉Co₇ very weak ferromagnetism and some form of superconductivity may coexist^{42,43}, but complex metallurgical properties, strong impurity scattering and ambiguities concerning the character of the magnetic electrons and the superconducting carriers have hampered progress in understanding. We note that measurements of the heat capacity, residual resistivity and upper critical field suggest that the carrier mean free path in Y₉Co₇ is considerably smaller than the superconducting coherence length. This would seem to preclude in this case the possibility of spin-triplet magnetically mediated superconductivity.

We conclude that an unequivocal example of magnetically mediated superconductivity connected with the simplest case of ferromagnetism, as opposed to the more complex case of antiferromagnetism, is still lacking. Here we report the results of a search for this kind of superconductivity in the low-temperature metallic-ferromagnet UGe₂, which appears to satisfy the various requirements mentioned above and which may not have the disadvantages of the systems described thus far. Our study may also test the alternative models of ferromagnetism and superconductivity of refs 44 and 45.

Band magnetism in UGe₂

UGe₂ crystallizes congruently from the melt into an orthorhombic structure with full inversion symmetry^{46,47}. Single crystals grow readily in the stoichiometric state, and specimens with unusually high purity can be prepared. In the best cases residual resistivities as

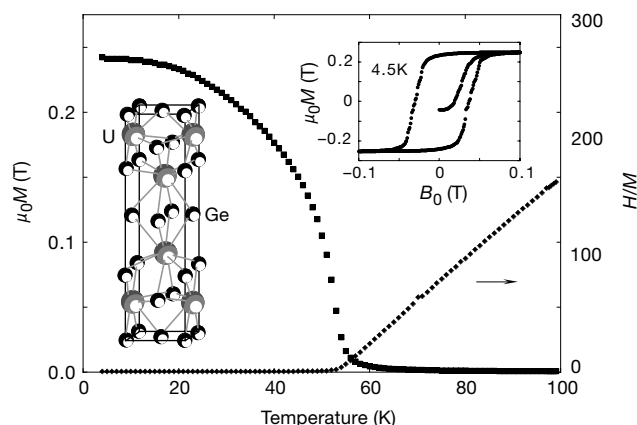


Figure 1 The magnetization and inverse magnetic susceptibility of UGe₂. Results (in SI units) are shown for a single crystal at ambient pressure along the easy axis in an applied field of 0.1 T. Left inset, the orthorhombic unit cell (the easy axis, or *a*-axis, is along the horizontal in the page); right inset, the typical form of the hysteresis loop at 4.5 K along the easy axis. The specimens used for magnetization and a.c. susceptibility measurements (Fig. 3) have demagnetizing factors along the easy axis of less than 0.2. *M*, magnetization; *H*, magnetic field; *B*₀, applied magnetic induction.

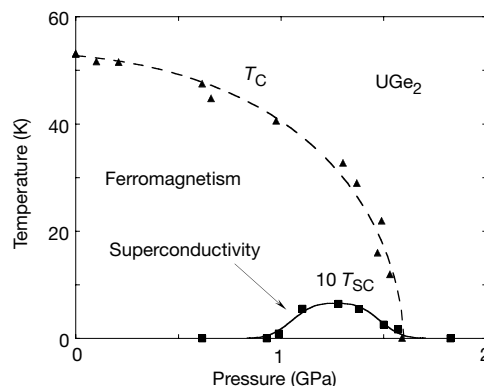


Figure 2 The temperature–pressure phase diagram of UGe₂. T_C denotes the Curie temperature and T_{SC} the superconducting transition temperature; the latter is determined from the 50% drop in resistivity, and the former from the cusp in the resistivity or the peak in the a.c. susceptibility (see for example, refs 50 and 53). (We note that the T_{SC} values are scaled by a factor of 10. For T_C versus pressure, see also ref. 49. The dashed and solid lines serve only to connect the data.)

low as $0.2 \mu\Omega \text{ cm}$ have been reported, corresponding in a conventional analysis to carrier mean free paths of several thousand ångströms (ref. 47). Ferromagnetic order is observed below a T_C that starts at 53 K at ambient pressure and falls monotonically towards absolute zero at a critical pressure p_c of 1.6 to 1.7 GPa, that is, at a pressure readily accessible with conventional piston and cylinder cells^{48–50}. The magnetic transition near p_c appears to be of first order⁵⁰, as was also noted earlier in the case of MnSi (ref. 35). This implies that the longitudinal susceptibility that enters the magnetic interaction potential cannot be tuned with pressure to arbitrarily high values at low temperatures. Thus, the conditions for magnetically mediated superconductivity need not be fulfilled at any pressure. As we shall see, however, superconductivity does indeed occur at least in the millikelvin range in a narrow pressure window in the ferromagnetic state just below p_c .

In the uranium compounds known as ‘heavy-ferrion systems’ the $5f$ levels are highly localized and, in conjunction with other more delocalized levels with which they couple, give rise to gapless ferrion excitations characterized by effective masses approaching (in extreme cases) that of protons or neutrons. In UGe_2 , however, the $5f$ electrons are more itinerant than in many heavy-ferrion systems, and Fermi-surface studies together with heat capacity measurements suggest that they behave more like the $3d$ electrons in the traditional itinerant-electron ferromagnets such as Fe, Co and Ni (refs 47, 51, 52). We stress that the magnetization in the latter ferromagnets arises from an exchange spin splitting of conduction electron bands, and not from a polarization of strictly localized electrons as is the case, for example, in the ferromagnetic metal Gd. UGe_2 differs from the $3d$ metals mainly in having a stronger spin-orbit interaction that leads to an unusually large magnetocrystalline anisotropy. The magnetization in UGe_2 is held along the easy magnetic axis by an anisotropy field of the order 100 T which is two or more orders of magnitude greater than in the $3d$ metals. As discussed in ref. 18, this anisotropy may be particularly favourable for magnetic pairing in the triplet channel. The important point is that, in the best available specimens, the f electrons give rise to fully itinerant quasiparticles with sufficiently long mean free paths to allow, at least at very low temperatures and near p_c , for the possibility of spin-triplet magnetically mediated superconductivity of the type considered here.

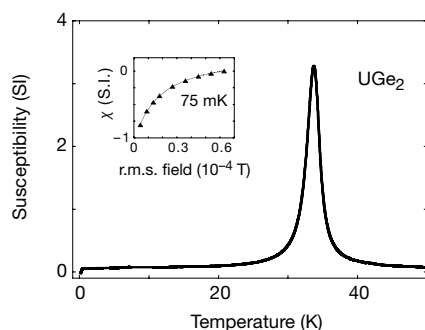


Figure 3 The a.c. susceptibility χ of UGe_2 at high pressure. The main figure shows the typical form of χ along the easy axis in the Earth’s magnetic field and at approximately 1.3 GPa. The rapid drop of the susceptibility upon entering the ferromagnetic state is consistent with the behaviour of the initial susceptibility observed in hysteresis measurements at ambient pressure (see Fig. 1 right inset). This behaviour is most naturally understood in terms of the abnormally strong magnetic anisotropy that characterizes this material. The interpretation of the observed drop of the susceptibility in the superconducting state (inset; at approximately 1.5 GPa) must take account of the presence of the internal field due to the ferromagnetic magnetization that is expected to lead to a spontaneous vortex lattice even in the absence of an external applied magnetic field^{43,62}. The existence of an internal field of the order of 0.1 to 0.2 T is consistent with our measurements of the upper critical field versus temperature.

Measurements of UGe_2

The magnetization and susceptibility versus temperature of UGe_2 at ambient pressure and in an applied field of 0.1 T are shown in Fig. 1, for the easy axis of magnetization. Also shown, as insets to this figure, are the unit cell and a typical hysteresis loop with the applied field along the easy axis. The form of the hysteresis curve and the difference in the magnetization along the easy and hard axes are consistent with an extremely strong magnetocrystalline anisotropy as discussed earlier. We note that one effect of this anisotropy is to greatly suppress at the centre of the hysteresis loop the initial susceptibility, which is governed by the way in which magnetic domains can grow or rotate. We also note that the data of Fig. 1 for the easy axis yield a moment per U atom of $\sim 1.4 \mu_B$ within a single domain at low temperatures and low fields, but an effective paramagnetic moment above T_C of $\sim 2.7 \mu_B$. This difference in values, and in particular a ratio of the paramagnetic to the ordered moment that is greater than unity, is widely observed in the $3d$ magnetic metals and is consistent with the picture of band or itinerant electron magnetism that is supported by Fermi-surface studies in UGe_2 .

The central result of our studies is the temperature–pressure phase diagram presented in Fig. 2. It shows that T_C falls monotonically with increasing pressure, drops precipitously above 1.5 GPa, and appears to vanish at a critical pressure p_c of between 1.6 to 1.7 GPa (ref. 50). Values of T_C determined from the resistivity, a.c. and d.c. magnetization measurements and elastic neutron scattering, are found to be consistent within experimental error. As in a previous study in MnSi, but now to a more extreme degree, the magnetic transition versus temperature becomes first order just below p_c but the critical temperature versus pressure nevertheless appears to remain continuous⁵⁰.

In a narrow range in pressure below p_c and thus within the ferromagnetic state, we observe a sudden and complete loss of resistivity in the millikelvin temperature range below T_{SC} (Fig. 2). The abrupt loss of resistivity just below p_c appears to be a robust property of UGe_2 , and has been observed in all of the eight samples that we have investigated. The survival of bulk ferromagnetism below T_{SC} has been confirmed directly via elastic neutron scattering measurements (A. Huxley *et al.*, manuscript in preparation). Also, we note that the a.c. susceptibility below T_{SC} tends to the limit of -1 (in SI units), as expected in the presence of bulk superconductivity (Fig. 3). The upper critical field B_{c2} is in excess of 3 T near the maximum of T_{SC} and is much higher than normally expected for

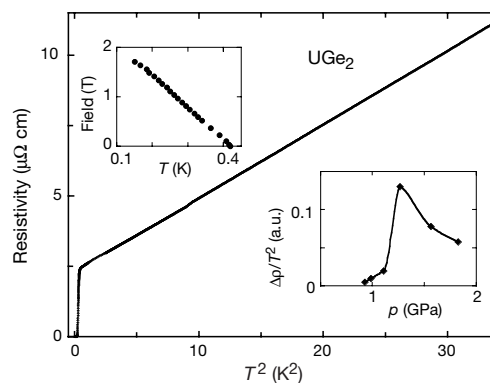


Figure 4 The resistivity ρ of UGe_2 at high pressure. The main figure shows our initial observation of a superconducting transition in ρ in an unaligned crystal of UGe_2 in the ferromagnetic regime at approximately 1.4 GPa. This plot of resistivity against the square of the temperature is roughly consistent with the form $\rho(T) = \rho_0 + AT^2$. The lower inset illustrates the typical pressure dependence of A around p_c and the upper inset gives an example of the variation of the superconducting transition temperature with external magnetic field. (For a related study of A versus p see also ref. 48. The solid line serves only to connect the data.)

conventional superconducting inclusions⁵³. Moreover, the apparent initial gradient $\partial H_C/\partial T$ near T_{SC} is anomalously large, and consistent with that expected in the presence of an internal field arising from ferromagnetic order.

As shown in Fig. 4 the resistivity $\rho(T)$ above T_{SC} , but typically below 10 K, is roughly described by an expression of the form $\rho(T) = \rho_0 + AT^2$, where ρ_0 is the residual resistivity and A is the quadratic coefficient. This is the form expected from the mutual scattering of fermion quasiparticles within the standard theory of metals; that is, within Fermi-liquid theory when the direct contribution to $\rho(T)$ from scattering of quasiparticles by well-defined collective modes such as phonons and magnons can be neglected. The parameters ρ_0 and A are not constants but vary with crystallographic direction and, more importantly, with pressure⁴⁸. As shown in the lower inset of Fig. 4, the quadratic coefficient A has a very pronounced pressure dependence and rises rapidly at a pressure $p_x \approx 1.2$ GPa close to (but distinctly below) p_c and within the narrow pressure window where superconductivity is indeed observed. We note that as p_x is approached, the range of validity of the above Fermi-liquid form of the resistivity collapses⁴⁸. The temperature variation of $\rho(T)$ is then more usefully described by a temperature exponent well below the Fermi liquid value of 2 (at low T but above T_{SC}).

Naively, we expect the quasiparticle interactions and thus A to be strongest at the critical boundary separating ferromagnetism and paramagnetism; namely, at p_c rather than at the lower pressure p_x , as observed. There is, however, evidence for the existence of a cross-over anomaly within the ferromagnetic state whose characteristic temperature T_x collapses towards zero also near p_x . This anomaly shows up typically as a weak and broad maximum in the temperature derivatives of the resistivity and the magnetization versus temperature⁴⁸. It also appears in the temperature dependence of the thermal expansion coefficient⁴⁹. We discuss below the possible origin of this cross-over phenomenon or transition, and its potential relevance to the pressure dependence of A and to superconductivity.

Role of magnetic interactions

We begin by comparing our findings with the predictions of the simplest model of magnetic interactions for an itinerant-electron ferromagnet. First, we consider more fully the facts that support the itinerant-electron picture, which is crucial to our discussion. Band structure analyses show that magneto-oscillatory phenomena observed in UGe₂ can be understood in terms of conventional energy bands calculated in the hypothetical paramagnetic state but with the Fermi level for majority and minority spins separated by an exchange splitting of the order of 70 meV (refs 47, 51). (This conclusion was based on an early determination of the crystal structure but is also expected to hold for the slightly revised crystal structure given in refs 54, 55.) The required exchange splitting is not large enough to entirely fill the majority spin bands; thus the Fermi surface consists of both majority and minority spin sheets. The masses of the quasiparticle excitations on these sheets of the Fermi surface are of the order of 20 times the bare electron mass or lower, and are consistent with the observed linear coefficient of the heat capacity of 35 mJ per mol U per K² (ref. 47). These findings, and the high ratio of the effective paramagnetic moment to the ordered moment discussed above, are generally consistent with the behaviour that has long been associated with the traditional itinerant-electron ferromagnets.

Within the simplest itinerant-electron model, the longitudinal magnetic susceptibility that enters the magnetic interaction potential for spin-triplet pairing is expected to have a maximum in the low-temperature limit around p_c . This should lead to a maximum in the resistivity coefficient A at p_c , as already noted, and spin-triplet magnetically mediated superconductivity over a narrow pressure range both above and below p_c . If the magnetic transition is strictly continuous or second order, a narrow dip in T_{SC} may arise very close

to p_c where pair formation tends to be frustrated by strong quasiparticle damping^{7,18}. This dip is expected to be suppressed, however, when the transition is first order. Moreover, when the transition is strongly first order, as is the case in UGe₂, the pressure range in which T_{SC} is finite is expected to contract and even collapse entirely on one or both sides of p_c .

The observation of superconductivity on just one side of p_c is not necessarily inconsistent with this model, but the shift in the peak in A versus pressure from p_c to the lower pressure p_x associated with the cross-over anomaly, would seem to require a more elaborate picture. A possible explanation for these features that is still broadly consistent with a spin-triplet magnetic interaction model is offered by a rather special property of the majority electron sheet of UGe₂. In the hypothetical paramagnetic state, the Fermi surface consists of an electron sheet having a quasi-cylindrical shape and a more complex hole sheet enclosing an equal volume to that of the electron sheet. In the presence of the exchange splitting, the minority-spin electron surface contracts greatly while the majority-spin counterpart expands towards the boundaries of the Brillouin zone and is cut off by the zone walls on two sides. What remains of the majority electron surface are two large and roughly parallel sheets reminiscent of a quasi-one-dimensional (quasi-1D) system⁵¹. Because the other sheets of the Fermi surface remain essentially three-dimensional, the resistivity and magnetic susceptibility are not expected to show quasi-1D behaviour. However, the 'hidden' quasi-1D character associated with the large and potentially important electron sheet of the majority Fermi surface can have profound consequences. In particular, the strong nesting of this sheet is expected to lead to very strong magnetic interactions between carriers of the same spin, peaked periodically in space at distances defined by the inverse of the nesting wave vector.

This quasi-1D model would tend to drive a transition from a simple ferromagnetic state to a modulated ferromagnetic state once a well defined exchange splitting, and hence quasi-1D character of the majority electron surface, has developed. This may provide a natural explanation of the cross-over anomaly T_x observed well below T_C and p_c . The peak in A at p_x instead of p_c would then be due to the fact that the ferromagnetic transition is strongly first order and magnetic interactions associated with it are relatively weak, while the lower transition (or a tendency) to a modulated structure is more nearly second order and hence produces strong magnetic interactions and pronounced quasiparticle scattering at low temperatures. As in the simplest model, this more elaborate scheme is also expected to lead naturally to spin-triplet magnetically mediated superconductivity. This is because the interactions are strong only between quasiparticles on the quasi-1D sheet of the Fermi surface which are all of the same spin, and also because the form of the magnetic interaction potential itself favours p -wave pairing. We stress that our model is not the same in detail as that applied to the quasi-1D organic compounds⁵⁶ in which the transport properties are highly anisotropic and metallic magnetism tends to be suppressed, nor is it the same as that traditionally applied to ³He (refs 2–6)—in which the Fermi surface is a simple sphere, and magnetic interactions are non-oscillatory in space.

For completeness we also consider briefly the viability of alternative available models for the superconductivity that we observe. Recently, the coexistence of ferromagnetism and spin-singlet s -wave superconductivity has been reconsidered^{44,45}. Singlet pairing cannot be ruled out when the exchange splitting of the Fermi surface is extremely small; for example, at the very border of a magnetic transition of second order at low T . In UGe₂, however, the ferromagnetic transition is strongly first order and the exchange splitting is large. Also, the proposed s -wave model does not seem to provide a natural interpretation of the cross-over anomaly, the pressure dependence of A , nor for the absence of superconductivity under similar conditions in the ferromagnetic states of several closely related metals. (We note that the deleterious effects of impurities or

breakdown of inversion symmetry that were discussed for odd parity or p -wave pairing do not extend in the same degree to s -wave pairing.)

The mechanisms that we have considered thus far are electronic in nature, and involve the lattice vibrations only in an indirect way—for example, in possibly modifying the phenomenological parameters that enter the quasiparticle interaction potential. A more direct role of the dynamics of the lattice for pairing in an itinerant-electron ferromagnet cannot be ruled out, but it would probably be considerably more exotic than in the traditional BCS model. In the band model appropriate to UGe_2 , pairing in the spin-singlet channel is unlikely because states of opposite momentum and of opposite spin on the exchange-split Fermi surface are only rarely (if ever) degenerate. At least in the presence of inversion symmetry, on the other hand, states of opposite momentum and of the same spin are always degenerate, and can readily be paired *en masse* in a zero-total-momentum state. Thus, it would seem that mechanisms for superconductivity based in some way more directly on lattice dynamics must also, as with the case of magnetic interactions, be compatible with spin-triplet odd-parity pairing.

As previously noted, the kind of superconductivity that we consider here can only arise if the carrier mean free path l exceeds the characteristic superconducting coherence length, ξ . From the measured upper critical field, we estimate in the standard way that ξ is of the order of 200 Å for pressures near p_x . This is indeed substantially lower than l , which is estimated from values of ρ_0 of our samples to be of the order of or greater than 1,000 Å (ref. 57). It is also of interest to examine whether the above order of magnitude of ξ is indeed reasonable within the itinerant-electron model we are considering. In the standard description, ξ depends on T_{SC} , the typical quasiparticle mass and the typical Fermi wavevector. The latter is not sensitive to pressure and can be estimated from zero pressure properties, but the quasiparticle mass is expected to vary strongly with pressure and is more difficult to determine. Within the magnetic interactions model the principal contribution to this mass is due to self interactions, which are naturally also connected to the quasiparticle scattering rate and hence to A . This leads to a simple connection between A and the linear coefficient of the heat capacity γ , which is proportional to the average of the quasiparticle mass. The connection can be expressed roughly as $A/\gamma^2 \approx 10 \mu\Omega \text{ cm K}^{-2}$ per $(\text{J mol}^{-1} \text{ K}^{-2})^2$, provided neither γ nor A are too close to singularities⁵⁸. This formula is found to be obeyed in order of magnitude by many metals on the border of magnetic order⁵⁹. It is also obeyed approximately by UGe_2 at ambient pressure, and from the known pressure variation of A , it suggests that the typical quasiparticle mass near p_x may be nearly two orders of magnitude greater than the bare electron mass. This leads us to a value of ξ in the itinerant model of a few hundred ångströms and thus of the same order of magnitude as that determined from critical field measurements. We note that at least near p_x , the estimated characteristic quasiparticle mass in UGe_2 is approaching in magnitude that of heavy-fermion metals such as UPt_3 (refs 21, 60). In contrast to UGe_2 , however, UPt_3 and the other known heavy-fermion superconductors are either paramagnetic or antiferromagnetic in their normal non-superconducting states. For this and the other reasons given above, UGe_2 appears to be unique among the known superconductors.

Spin-triplet superconductivity

We have observed superconductivity in the ferromagnetic state of the itinerant-electron system UGe_2 . Superconductivity exists only within a narrow window in lattice density, close to a critical density where ferromagnetism disappears. Our findings have been discussed in relation to a quasiparticle interaction model that favours a spin-triplet magnetically mediated form of superconductivity. This differs from the simplest model of magnetic pairing in a ferromagnet in that it includes the effects of a quasi-1D majority-spin

sheet of the Fermi surface embedded in an otherwise three-dimensional Fermi surface that characterizes this material. The effects of nesting of this majority-spin sheet can account qualitatively for the observed cross-over anomaly, and leads naturally to magnetic pairing in the spin-triplet channel in a narrow range of density within the ferromagnetic side of the critical density. Because the ferromagnetic state out of which superconductivity arises is well described within a band model, even more exotic models than the one we propose here would seem to require spin-triplet pairing. UGe_2 provides us with a new, and apparently unique, example of superconductivity in the field of magnetism.

Note added in proof: Additional theoretical analyses relevant to this work may be found in refs 63 and 64. For a more complete discussion of the background of this research and of our initial search for magnetically mediated superconductivity in UGe_2 see ref. 65. □

Methods

Samples of UGe_2 were grown by radio-frequency induction melting in water-cooled copper crucibles in ultrahigh-vacuum chambers. Details of the preparation techniques will be given elsewhere. The crystals selected for investigation have residual resistivities as low as a few tenths of $1 \mu\Omega \text{ cm}$. Also, no secondary chemical phases have been detected by electron-probe microanalysis with a resolution of 100 nm and accuracy of 1% of composition.

Measurements of the electrical resistivity and a.c. magnetic susceptibility in both aligned crystals and polycrystalline samples have been performed as a function of temperature, magnetic field and pressure. The pressure cells are of the piston-cylinder type made of non-magnetic alloys suitable for efficient cooling to the low millikelvin range and for the application of external magnetic fields⁶¹. The pressure-transmitting fluids employed within these cells are those found to produce hydrostatic pressures of adequate homogeneity for sensitive quantum oscillation studies, for example. The low millikelvin range was reached by means of a dilution refrigerator (Oxford Instruments) or a demagnetization refrigerator (Cambridge Magnetic Refrigeration) with base temperatures of 20 mK and 100 mK, respectively, in the presence of a standard pressure cell. Studies as a function of magnetic field were made with a 6-T superconducting magnet attached to the dilution refrigerator system.

The resistivity was measured by the conventional four-terminal method using common mode voltage rejection circuitry and low-temperature transformers to allow adequate voltage sensitivity with the very low excitation powers needed for acceptable levels of sample heating. The a.c. magnetic susceptibility was studied with similar electronics and in zero d.c. external magnetic field by means of a miniature niobium modulation coil wound on top of a compensated copper pick-up coil mounted as an assembly inside the pressure cells. The sample temperature was measured in the millikelvin range by means of calibrated RuO_2 or carbon resistance thermometers thermally anchored to the outside of the Au-plated pressure cells and also anchored to the samples directly via copper leads passing into the pressure cell. To ensure that the samples and thermometers were at the same temperature, very slow rates of sweep were employed (typically 1 mK min^{-1} below 1 K), and it was checked that up and down temperature ramps gave equivalent results independent of the excitation power in the ranges of interest.

Finally, bulk magnetization measurements as a function of magnetic field and temperature were performed by means of a SQUID magnetometer (Quantum Design) with a base temperature of 2 K. A miniature pressure cell only 8 mm in diameter was constructed out of high-purity non-magnetic BeCu to allow high-precision measurements to be extended from atmospheric pressure to applied pressures of up to approximately 1.5 GPa. Elastic neutron scattering measurements have also been carried out in this temperature and pressure range. Details of these measurements will be given elsewhere.

Received 19 May; accepted 27 June 2000.

- Bardeen, J., Cooper, L. N. & Schrieffer, J. R. Theory of superconductivity. *Phys. Rev.* **108**, 1175–1204 (1957).
- Brueckner, K. A., Soda, T., Anderson, P. W. & Morel, P. Level structure of nuclear matter and liquid ^3He . *Phys. Rev.* **118**, 1442–1446 (1960).
- Nakajima, S. Paramagnon effect on the BCS transition in ^3He . *Prog. Theor. Phys.* **50**, 1101–1109 (1973).
- Brinkman, W. F., Serene, J. W. & Anderson, P. W. Spin-fluctuation stabilization of anisotropic superfluid states. *Phys. Rev. A* **10**, 2386–2394 (1974).
- Leggett, A. J. A theoretical description of the new phases of liquid ^3He . *Rev. Mod. Phys.* **47**, 331–414 (1975).
- Levin, R. & Valls, O. T. Strong-coupling theory of superfluid transition temperatures for paramagnon models: application to ^3He . *Phys. Rev. B* **17**, 191–200 (1978).
- Fey, D. & Appel, J. Coexistence of p -state superconductivity and itinerant ferromagnetism. *Phys. Rev. B* **22**, 3173–3182 (1980).
- Hirsch, J. E. Attractive interaction and pairing in fermion systems with strong on-site repulsion. *Phys. Rev. Lett.* **54**, 1317–1320 (1985).
- Miyake, K., Schmitt-Rink, S. & Varma, C. M. Spin-fluctuation mediated even-parity pairing in heavy-fermion superconductors. *Phys. Rev. B* **34**, 6554–6556 (1986).
- Scalapino, D. J., Loh, E. Jr, Hirsch, J. E. d -wave pairing near a spin-density-wave instability. *Phys. Rev. B* **34**, 8190–8192 (1986).
- Millis, A. J., Sachdev, S. & Varma, C. M. Inelastic-scattering and pair breaking in anisotropic and isotropic superconductors. *Phys. Rev. B* **37**, 4975–4986 (1988).

12. Bickers, N. E., Scalapino, D. J. & White, S. R. Conserving approximations for strongly correlated electron-systems - Bethe-Salpeter equation and dynamics for the two-dimensional Hubbard-model. *Phys. Rev. Lett.* **62**, 961–964 (1989).
13. Pines, D. Spin excitations and superconductivity in cuprate oxide and heavy electron superconductors. *Physica B* **163**, 78–88 (1990).
14. Moriya, T., Takahashi, Y. & Ueda, K. Antiferromagnetic spin fluctuations and superconductivity in 2-dimensional metals - a possible model for high- T_c oxides. *J. Phys. Soc. Jpn* **52**, 2905–2915 (1990).
15. Monthoux, P., Balatsky, A. V. & Pines, D. Towards a theory of high temperature superconductivity in the antiferromagnetically correlated cuprate oxide. *Phys. Rev. Lett.* **67**, 3448–3451 (1991).
16. Bulut, N., Hone, D. W., Scalapino, D. J. & Bickers, N. E. Knight-shifts and nuclear-spin-relaxation rates for 2-dimensional models of CuO_2 . *Phys. Rev. B* **41**, 1797–1811 (1990).
17. Schrieffer, J. R., Wen, X. G. & Zhang, S. C. Spin-bag mechanism of high-temperature superconductivity. *Phys. Rev. Lett.* **60**, 944–947 (1988).
18. Monthoux, P. & Lonzarich, G. G. p -wave and d -wave superconductivity in quasi-two-dimensional metals. *Phys. Rev. B* **59**, 14598–14605 (1999).
19. Steglich, F. *et al.* Superconductivity in the presence of strong Pauli paramagnetism: CeCu_2Si_2 . *Phys. Rev. Lett.* **43**, 1892–1896 (1979).
20. Ott, H. R., Rudigier, H., Fisk, Z. & Smith, J. L. UBe_{13} —an unconventional actinide superconductor. *Phys. Rev. Lett.* **50**, 1595–1598 (1983).
21. Stewart, G. R. Heavy-fermion systems. *Rev. Mod. Phys.* **56**, 755–787 (1984).
22. Fisk, Z. *et al.* Heavy-electron metals: new highly correlated states of matter. *Science* **239**, 33–42 (1988).
23. Jerome, D. *Organic Conductors* (Dekker, New York, 1994).
24. Ishiguro, T. & Yamaji, K. (eds) *Organic Conductors* Ch. 3 (Springer, Berlin, 1990).
25. Jaccard, D., Behnia, K. & Sierro, J. Pressure-induced heavy fermion superconductivity of CeCu_2Ge_2 . *Phys. Lett. A* **163**, 475–480 (1992).
26. Grosche, F. M., Julian, S. R., Mathur, N. D. & Lonzarich, G. G. Magnetic and superconducting phases in CePd_2Si_2 . *Physica B* **224**, 50–52 (1996).
27. Movshovich, R. *et al.* Superconductivity in heavy-fermion CeRh_2Si_2 . *Phys. Rev. B* **53**, 8241–8244 (1996).
28. Julian, S. R. *et al.* The normal states of magnetic d and f transition metals. *J. Phys. Condens. Matter* **8**, 9675–9688 (1996).
29. Walker, I. R., Grosche, F. M., Freye, D. M. & Lonzarich, G. G. The normal and superconducting states of CeIn_3 near the border of antiferromagnetic order. *Physica C* **282–287**, 303–306 (1997).
30. Fukuyama, H. Electronic phase transition tuned by pressure: superconductivity and antiferromagnetism. *Rev. High Pressure Sci. Technol.* **7**, 465–468 (1998).
31. Mathur, N. D. *et al.* Magnetically mediated superconductivity in heavy fermion compounds. *Nature* **394**, 39–43 (1998).
32. Jourdan, M., Huth, M. & Adrian, H. Superconductivity mediated by spin fluctuations in the heavy-fermion compound UPd_2Al_3 . *Nature* **398**, 47–49 (1999).
33. Lonzarich, G. G. in *Electron* (ed. Springford, M.) Ch. 6 (Cambridge Univ. Press, Cambridge, 1997).
34. Grosche, F. M., Pfleiderer, C., McMullan, G. J., Lonzarich, G. G. & Bernhoeft, N. R. Critical behaviour of ZrZn_2 . *Physica B* **206 & 207**, 20–22 (1995).
35. Pfleiderer, C., McMullan, G. J., Julian, S. R. & Lonzarich, G. G. Magnetic quantum phase transition in MnSi under hydrostatic pressure. *Phys. Rev. B* **55**, 8330–8338 (1997).
36. Thessieu, C., Pfleiderer, C. & Flouquet, J. Thermodynamical study under hydrostatic pressure of MnSi . *Physica B* **239**, 67–70 (1997).
37. Maeno, Y. *et al.* Superconductivity in a layered perovskite without copper. *Nature* **372**, 532–534 (1994).
38. Rice, T. M. An analogue of superfluid ^3He . *Nature* **396**, 627–629 (1998).
39. Sidis, Y. *et al.* Evidence for incommensurate spin fluctuations in Sr_2RuO_4 . *Phys. Rev. Lett.* **83**, 3320–3323 (1999).
40. Tallon, J. *et al.* Coexisting ferromagnetism and superconductivity in hybrid rutheno-cuprate superconductors. *IEEE Trans. Appl. Supercond.* **9**, 1696–1699 (1999).
41. Pickett, W. E., Weht, R. & Shick, A. B. Superconductivity in ferromagnetic $\text{RuSr}_2\text{GdCu}_2\text{O}_8$. *Phys. Rev. Lett.* **83**, 3713–3716 (1999).
42. Kolodziejczyk, A., Sarkissian, B. V. B. & Coles, B. R. Magnetism and superconductivity in a transition metal compound: Y_4Co_3 . *J. Phys. F* **10**, L333–L337 (1980).
43. Sinham K. P. & Kakani, S. L. *Magnetic Superconductors: Recent Developments* (Nova Science, New York, 1989).
44. Blagoev, K. B., Engelbrecht, J. R. & Bedell, K. S. Effect of ferromagnetic spin correlations on superconductivity in ferromagnetic metals. *Phys. Rev. Lett.* **82**, 133–136 (1999).
45. Krachev, N. I., Blagoev, K. B., Bedell, K. S. & Littlewood, P. B. Coexistence of superconductivity and ferromagnetism in ferromagnetic metals. Preprint cond-mat/9911489 at (<http://xxx.lanl.gov>) (1999; cited 30 Nov. 1999).
46. Menovsky, A., de Boer, F. R., Frings, P. H. & Franse, J. J. M. in *High Field Magnetism* 189 (ed. Date, M.) (North-Holland, Amsterdam, 1983).
47. Satoh, K. *et al.* de Haas-van Alphen effect in UGe_2 . *J. Phys. Soc. Jpn* **61**, 1827–1828 (1992).
48. Oomi, G., Kagayama, T. & Onuki, Y. Critical electron scattering in UGe_2 near the magnetic phase transition induced by pressure. *J. Alloys Compounds* **271–273**, 482–485 (1998).
49. Nishimura, K., Oomi, G., Yun, S. W. & Onuki, Y. Effect of pressure on the Curie temperature of single-crystal UGe_2 . *J. Alloys Compounds* **213**, 383–386 (1994).
50. Huxley, A., Sheikin, I. & Braithwaite, D. Metamagnetic behaviour near the quantum critical point in UGe_2 . *Physica B* **284 & 288**, 1277–1278 (2000).
51. Yamagami, H. & Hasegawa, A. Fermi surface of the ferromagnetic heavy-electron compound UGe_2 . *Physica B* **186–188**, 182–184 (1993).
52. Lonzarich, G. G. Band structure and magnetic fluctuations in ferromagnetic or nearly ferromagnetic metals. *J. Magn. Magn. Mater.* **45**, 43–53 (1984).
53. Agarwal, P. Magnetism and superconductivity in heavy fermion metals. Thesis, Univ. Cambridge (2000).
54. Oikawa, K., Kamiyama, T., Asano, H., Onuki, Y. & Kohgi, M., Crystal structure of UGe_2 . *J. Phys. Soc. Jpn* **65**, 3229–3232 (1996).
55. Boulet, P. *et al.* Crystal and magnetic structure of the uranium digermanide UGe_2 . *J. Alloys Compounds* **247**, 104–108 (1997).
56. Lee, I. J., Naughton, M. J., Danner, G. M. & Chaikin, P. M. Anisotropy of the upper critical field in $\text{TMTSF}_2\text{PF}_6$. *Phys. Rev. Lett.* **78**, 3555–3558 (1997).
57. Mackenzie, A. P. *et al.* Extremely strong dependence of superconductivity on disorder in Sr_2RuO_4 . *Phys. Rev. Lett.* **80**, 161–164 (1998).
58. Takimoto, T. & Moriya, T. Relationship between resistivity and specific heat in heavy electron systems. *Solid State Commun.* **99**, 457–460 (1996).
59. Kadowaki, K. & Woods, S. B. Universal relationship of the resistivity and specific heat in heavy-fermion compounds. *Solid State Commun.* **58**, 507–509 (1986).
60. Taillefer, L. & Lonzarich, G. G. Heavy-fermion quasiparticles in UPt_3 . *Phys. Rev. Lett.* **60**, 1570–1573 (1988).
61. Walker, I. R. Nonmagnetic piston-cylinder pressure cell for use at 35 kbar and above. *Rev. Sci. Instrum.* **70**, 3402–3412 (1999).
62. Bulaevskii, L. N., Buzdin, A. I., Panjukov, S. V. & Kulic, M. L. Coexistence of superconductivity and magnetism: theoretical predictions and experimental results. *Adv. Phys.* **39**, 175 (1985).
63. Rousseau, R. & Millis, A. J. Quantum critical effects on transition temperature of magnetically mediated p -wave superconductivity. Preprint cond-mat/0006208 at (<http://xxx.lanl.gov>) (2000; cited 26 May 2000).
64. Ohmi, T. & Machida, K. Nonunitary superconducting state in UPt_3 . *Phys. Rev. Lett.* **71**, 625–628 (1993).
65. Saxena, S. S. Magnetic and superconducting phases of heavy fermion compounds. Thesis, Univ. Cambridge (1998).

Acknowledgements

We thank in particular S. V. Brown and also F. Beckers, K. S. Bedell, K. B. Blagoev, D. M. Broun, P. Coleman, D. Forsythe, C. D. Frost, D. E. Khmel'nitskii, P. B. Littlewood, A. J. Millis, P. Niklowitz, T. T. M. Palstra, D. Pines, C. Pfleiderer, K. Sandeman, A. J. Schofield and A. Tselik for discussions. The work was supported in part by the Cambridge Research Centre in Superconductivity, the UK EPSRC, the Paul Instrument Fund of the Royal Society, the Cambridge Newton Trust and the Commonwealth Scholarship Commission. The work performed in Grenoble was supported by the CEA Direction des Sciences de la Matière.

Correspondence and requests for materials should be addressed to G.G.L. (e-mail: ltip-secretary@phy.cam.ac.uk).



Regulation of chromatin structure by site-specific histone H3 methyltransferases

Stephen Rea*, Frank Eisenhaber*, Dónal O'Carroll*, Brian D. Strahl†, Zu-Wen Sun†, Manfred Schmid*, Susanne Opravil*, Karl Mechtler*, Chris P. Ponting‡, C. David Allis† & Thomas Jenuwein*

* Research Institute of Molecular Pathology (IMP), The Vienna Biocenter, Dr. Bohrgasse 7, A-1030 Vienna, Austria

† Department of Biochemistry and Molecular Genetics, University of Virginia Health Science Center, Charlottesville, Virginia 22908, USA

‡ MRC, Functional Genetics Unit, Department of Human Anatomy and Genetics, University of Oxford, OX1 3QX, UK

The organization of chromatin into higher-order structures influences chromosome function and epigenetic gene regulation. Higher-order chromatin has been proposed to be nucleated by the covalent modification of histone tails and the subsequent establishment of chromosomal subdomains by non-histone modifier factors. Here we show that human *SUV39H1* and murine *Suv39h1*—mammalian homologues of *Drosophila Su(var)3-9* and of *Schizosaccharomyces pombe clr4*—encode histone H3-specific methyltransferases that selectively methylate lysine 9 of the amino terminus of histone H3 *in vitro*. We mapped the catalytic motif to the evolutionarily conserved SET domain, which requires adjacent cysteine-rich regions to confer histone methyltransferase activity. Methylation of lysine 9 interferes with phosphorylation of serine 10, but is also influenced by pre-existing modifications in the amino terminus of H3. *In vivo*, deregulated *SUV39H1* or disrupted *Suv39h* activity modulate H3 serine 10 phosphorylation in native chromatin and induce aberrant mitotic divisions. Our data reveal a functional interdependence of site-specific H3 tail modifications and suggest a dynamic mechanism for the regulation of higher-order chromatin.

Higher-order chromatin is essential for epigenetic gene control and for the functional organization of chromosomes. Differences in higher-order chromatin structure have been linked with distinct covalent modifications of histone tails that regulate transcriptional 'on' or 'off' states^{1–3} and influence chromosome condensation and segregation^{4,5}. Post-translational modifications of histone N termini, particularly of H4 and H3, are well documented and have functionally been characterized as changes in acetylation^{1–3}, phosphorylation⁵ and, most recently, methylation^{6,7}. In contrast to the large number of described histone acetyltransferases (HATs) and histone deacetylases (HDACs), genes encoding enzymes that regulate phosphorylation^{8,9} or methylation⁶ of histone N termini are only now being identified. Moreover, the interdependence of the different histone tail modifications for the integration of transcriptional output or higher-order chromatin organization is currently not understood.

Genetic screens for suppressors of position effect variegation (PEV) in *Drosophila*¹⁰ and *S. pombe*¹¹ have identified a subfamily of about 30–40 loci, which are referred to as *Su(var)*-group¹² genes. Several histone deacetylases^{13,14}, protein phosphatase type 1 (ref. 15) and S-adenosyl methionine synthetase¹⁶ have been classified as belonging to the *Su(var)* group. In contrast, *Su(var)2-5* (which is allelic to *HPI*)¹⁷, *Su(var)3-7* (ref. 18) and *Su(var)3-9* (refs 19, 20) encode heterochromatin-associated proteins. *Su(var)* gene function thus suggests a model, in which modifications at the nucleosomal level may initiate the formation of defined chromosomal subdomains that are then stabilized and propagated by heterochromatic SU(VAR) proteins²¹.

Su(var)3-9 is dominant over most PEV modifier mutations¹⁹, and mutants in the corresponding *S. pombe clr4* gene²², disrupt heterochromatin association of other modifying factors and result in chromosome segregation defects²³. We isolated *Su(var)3-9* homologues from humans (*SUV39H1*)²⁴ and mice (*Suv39h1*)²⁴, and showed that they encode heterochromatic proteins that associate with mammalian HP1 (ref. 24). The SU(VAR)3-9 protein family combines two of the most evolutionarily conserved domains

of 'chromatin regulators': the chromo^{25,26} and SET²⁷ domains. Whereas the 60-amino-acid chromo domain represents an ancient histone-like fold²⁸ that directs euchromatic or heterochromatic localizations²⁹, the molecular role of the 130-amino-acid SET domain has remained unknown. Here we report that mammalian SUV39H1 or Suv39h proteins are SET-domain-dependent, H3-specific histone methyltransferases (HMTases) which selectively methylate lysine 9 (Lys 9) of the H3 N terminus—a modification that appears intrinsically linked to the organization of higher-order chromatin.

The SET domain and methyltransferases

The evolutionarily conserved SET domain was initially characterized as a common motif in the PEV modifier SU(VAR)3-9 (ref. 19), the *Polycomb*-group protein E(Z)³⁰ and the *trithorax*-group protein TRX³¹. In addition to these founding *Drosophila* members, we detected SET domain motifs in over 140 gene sequences³² of diverse and even antagonistic functions ranging from *Saccharomyces cerevisiae* to man, but also including some bacteria and viruses. Using the SET domains of the SU(VAR)3-9 protein family as a starting alignment, we identified significant sequence and secondary-structure similarities to six plant methyltransferases (Fig. 1). Although some of these plant sequences have been classified as potential histone lysine N-methyltransferases, only one has been functionally characterized and was found to lack HMTase activity^{33,34}. All six plant sequences contain an insertion of about 100 amino acids in the middle of the SET domain and do not comprise a SET-domain-associated cysteine-rich region or the carboxy-terminal tail with its three conserved cysteines. Whereas the secondary structure and sequence similarities extend over the entire SET domain (data not shown), highest identities are found in two short amino-acid stretches in the C-terminal half of the SET domain core: NHSC and GE(x)₅Y (see Fig. 1). Our alignment suggests a bipartite structure of the SET domain, in which these two conserved motifs may be of particular functional importance.

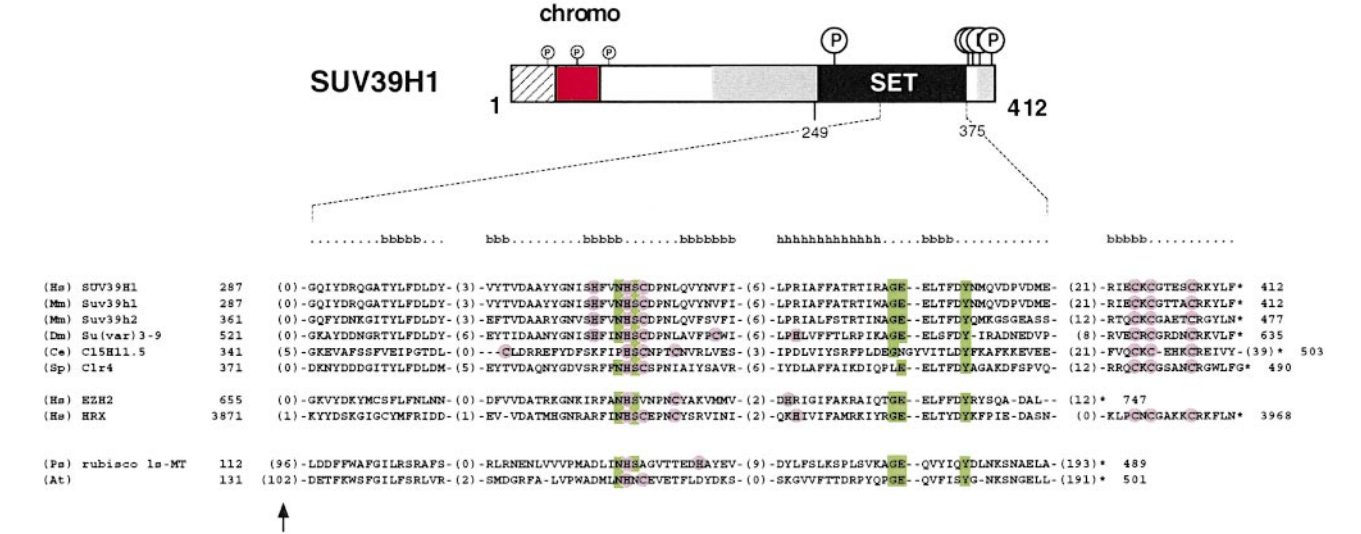


Figure 1 Sequence similarity of SET domains with plant methyltransferases. Top, the 412 amino acids of human SUV39H1 protein, highlighting the chromo domain (red box), the C-terminal SET domain (black box; residues 249–375) and SET domain-associated cysteine-rich regions (grey shading)^{24,35}. Bottom, amino-acid and secondary structure, β -sheet (b) or α -helix (h), similarities of the C-terminal halves of SET domain sequences from human SUV39H1 (ref. 24), murine Suv39h1 (ref. 24), murine Suv39h2 (D. O’C. *et al.*, manuscript in preparation), *Drosophila* SU(VAR)3-9 (ref. 19), a *C. elegans* SU(VAR)3-9-like

open reading frame C15H11.5 (CAB02737), *S. pombe* CLR4 (ref. 22), human EZH2 (ref. 36), the human trithorax homologue HRX³⁷ and MTases from *P. sativum* (rubisco ls-MT)^{33,34} and *A. thaliana* (O65218). The plant MTase sequences contain an insertion of about 100 amino acids in the middle of the SET domain (arrow). Cysteine and histidine residues are highlighted in pink, and identical amino-acid positions in at least 9 of the 10 SET domains are highlighted in green.

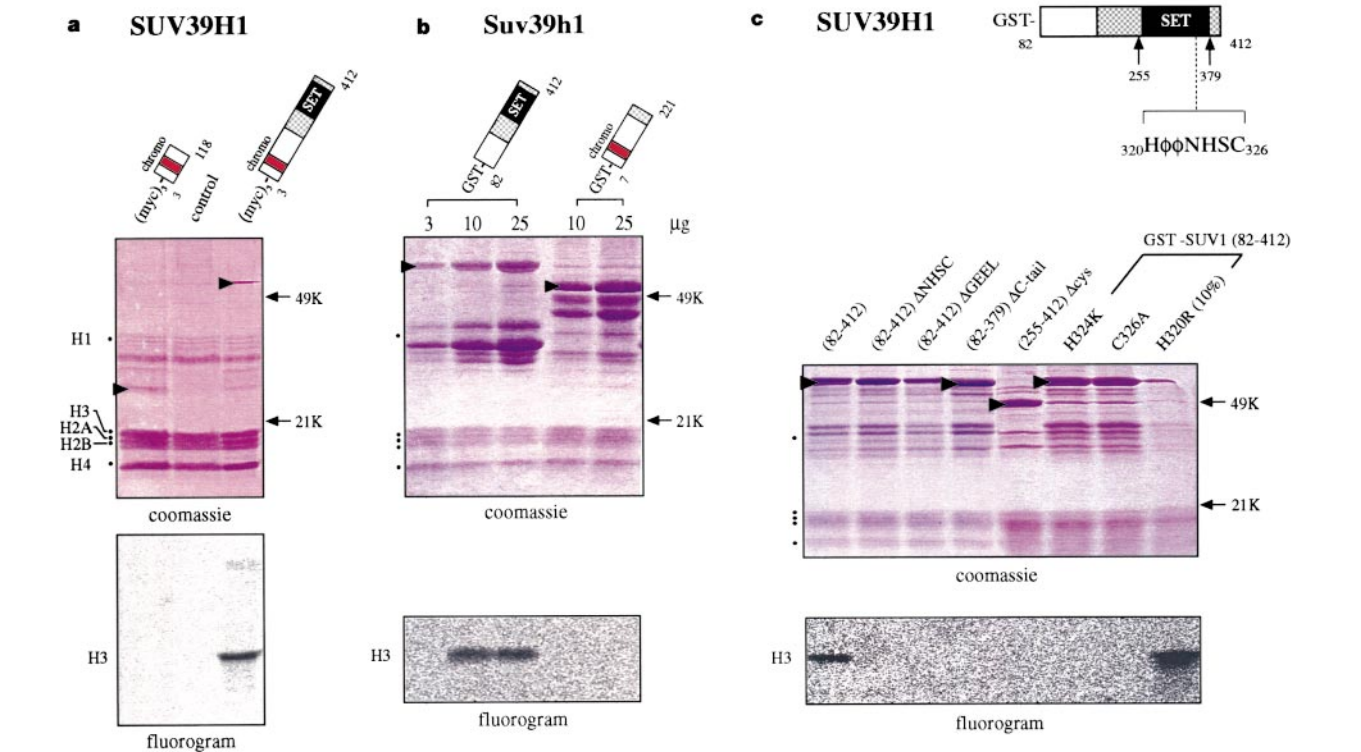


Figure 2 Histone methyltransferase activity of transfected and recombinant SUV39H1 and Suv39h1 proteins. **a**, Triple myc-tagged full-length human SUV39H1 (residues 3–412) or a C-terminally truncated SUV39H1 protein (residues 3–118) were immunoprecipitated from ‘stably’ transfected HeLa cell lines³⁵ and used in *in vitro* HMTase reactions with free histones as substrates and S-adenosyl-[methyl-¹⁴C]-L-methionine as methyl donor. Coomassie stain (top panel) shows purified proteins (arrowheads) and free histones (dots). Fluorography (bottom panel) indicates H3 HMTase activity of (myc)₃-SUV39H1.

b, Recombinant GST fusion proteins encoding different domains of murine Suv39h1 were used in increasing protein concentrations for *in vitro* HMTase reactions with free histones. **c**, About 10 μ g of the indicated fusion proteins encoding GST-SUV1 (82–412) (human SUV39H1) and seven GST-SUV39H1 mutants were analysed in *in vitro* HMTase reactions with free histones. For the hyperactive H320R mutant, only 1 μ g (10%) of the corresponding fusion product was used.

SUV39H1 and Suv39h1 possess HMTase activity

To investigate whether the SET domain of human SUV39H1 has enzymatic activity, we tested histones as possible substrates for *in vitro* methylation. Using HeLa cell lines 'stably' expressing³⁵ triple myc-tagged full-length SUV39H1 (residues 3–412), we enriched the ectopic protein from nuclear extracts by immunoprecipitation with anti-myc beads (Fig. 2a, top panel, arrowhead) and probed for activity to transfer a labelled methyl group from *S*-adenosyl-[methyl-¹⁴C]-L-methionine to free histones according to described conditions⁷. Reaction products were separated by polyacrylamide gel electrophoresis with SDS (SDS-PAGE) and visualized by fluorography, indicating selective transfer of the methyl label to H3 (see below). In contrast, no signals were detected with extracts from a HeLa cell line that expresses only the N-terminal third of SUV39H1 (residues 3–118) or with extracts from HeLa control cells.

To confirm that the HMTase activity is an intrinsic property of SUV39H1 and not mediated by a SUV39H1-associated factor, we repeated the *in vitro* HMTase reactions with recombinant products that were purified as glutathione-S-transferase (GST) fusion proteins from *Escherichia coli* (Fig. 2b, top panel, arrowheads). For this analysis, we used murine Suv39h1, which is 95% identical to human SUV39H1 (ref. 24). A purified GST product comprising residues 82–412 maintained HMTase activity (although at a reduced level as compared with transfected SUV39H1), whereas a purified GST product comprising residues 7–221 proved negative, even at higher protein concentrations (Fig. 2b, lower panel).

A catalytic motif in the SET domain

Similar to the recombinant murine GST–Suv1(82–412) product, the corresponding human SUV39H1 fusion protein, GST–SUV1(82–412), is catalytically active (Fig. 2c). We introduced short internal deletions (Δ NHSCDPN_{323–329}; Δ NHSC and Δ GEELTFDY_{358–365}; Δ GEEL) into the two conserved regions of the SET domain core in GST–SUV1(82–412), and also generated mutants that lack the C-terminal tail (Δ C-tail) or the SET-associated cysteine-rich region (Δ cys). None of the mutant proteins showed HMTase activity.

These results suggest that the cysteine-rich regions have a significant contribution, although their apparent absence in the plant methyltransferases does not prevent catalytic activity. To investigate enzyme function of the SET domain in more detail, we therefore introduced point mutations into the most highly conserved ³²⁰H ϕ ϕ NHSC₃₂₆ motif (where ϕ are hydrophobic residues) of GST–SUV1(82–412). In particular, we mutated the invariant histidine 324 to leucine (H324L; data not shown) or lysine (H324K), and also changed cysteine 326 to alanine (C326A) or replaced histidine 320 by arginine (H320R). *In vitro* HMTase assays indicated that all point mutations, with the exception of H320R, abolished enzymatic activity. Unexpectedly, the H320R mutation resulted in an hyperactive enzyme with activity increased 20-fold or more (Fig. 2c). These data define the ³²⁰H ϕ ϕ NHSC₃₂₆ motif in the SET domain as an important catalytic site.

HMTase activity within the SU(VAR)3-9 family

Because the SET domain is one of the most conserved protein motifs in chromatin regulators^{27,31}, we analysed whether SU(VAR)3-9 family members or other SET domain proteins contain HMTase activity. We generated GST fusion products of the extended SET domains of *S. pombe* CLR4 (ref. 22), human EZH2 (ref. 36) and human HRX³⁷ that would correspond to GST–SUV1(82–412) (Fig. 3a). GST–CLR4(127–490) displayed pronounced HMTase activity at 3–5-fold increased levels (Fig. 3b) compared with the recombinant SUV39H1 product, consistent with CLR4 carrying an arginine at the hyperactive position (see Fig. 1). In contrast, both GST–EZH2(382–747) and GST–HRX(3643–3966) had undetectable HMTase activity towards free histones (Fig. 3b), whereas a comparable GST product generated from the murine *Suv39h2* gene (D. O'C.

et al., manuscript in preparation), GST–Suv2(157–477), was as active as GST–SUV1(82–412). EZH2 lacks the C-terminal cysteines, and HRX does not contain the SET domain-associated cysteine-rich region. Both of these cysteine domains are present in CLR4, Suv39h2 and SUV39H1. In agreement with the mutational analysis of SUV39H1, it thus appears that HMTase activity towards free histones requires the combination of the SET domain with adjacent cysteine-rich regions, and is a quality found in only a restricted number of proteins containing SET domains.

H3 Lys 9 is the main site for the Suv39h1 HMTase

The above experiments indicated that the HMTase activity of mammalian SU(VAR)3-9 related proteins is selective for H3 under our assay conditions. To examine this finding in more detail, we carried out *in vitro* methylation reactions with individual histones, using GST–Suv1(82–412) as enzyme. H3 is specifically methylated, whereas no signals are detected with H2A, H2B or H4 (Fig. 4a). A weak signal is present if H1 was used as the sole substrate; the significance of H1 methylation remains to be determined.

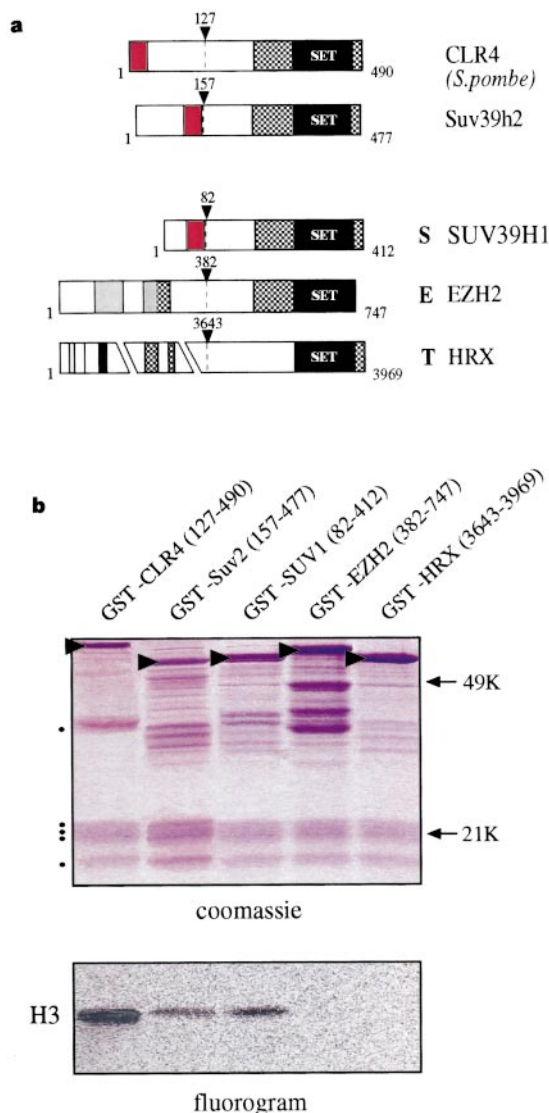


Figure 3 Specific HMTase activity of SU(VAR)3-9 related proteins. **a**, Diagram representing the domain structures of *S. pombe* CLR4, murine Suv39h2, human SUV39H1, human EZH2 and human HRX proteins. Arrowheads indicate the N-terminal fusion to GST. Cysteine-rich regions are indicated by grey stippling. **b**, About 10 μ g of the indicated GST fusion proteins were analysed in *in vitro* HMTase reactions with free histones. Coomassie stain (top panel) shows purified proteins (arrowheads) and free histones (dots). The murine Suv39h2 product is abbreviated as GST–Suv2(157–477).

Methylation of H3 occurs predominantly at Lys 4 in a wide range of organisms, as well as at Lys 9 in HeLa cells, although the responsible HMTase(s) have not been defined⁷. To investigate the methylation-site profile of Suv39h1, we tested unmodified peptides as substrates encoding the wild-type H3 N terminus (residues 1–20) and a mutant K9L peptide, changing Lys 9 to leucine. In addition, we included insulin and peptides encoding the N termini of CENP-A³⁸ and macroH2A³⁹. Peptides were methylated *in vitro* by GST–Suv1(82–412), and reaction products were separated by high percentage SDS–PAGE and visualized by fluorography. These *in vitro* assays revealed selective methylation of the wild-type H3 peptide, whereas no signals were detected with the CENP-A or macroH2A peptides, or with insulin (Fig. 4b). Notably, the mutated H3 (K9L) peptide was not a substrate, suggesting that Lys 9 of the H3 N terminus is a preferred residue for Suv39h1-dependent HMTase activity.

To determine more definitively this site preference, the wild-type H3 N-terminal peptide was methylated *in vitro* by GST–Suv1(82–

412), using S-adenosyl-[methyl-³H]-L-methionine. The labelled peptide, purified by reverse-phase high performance liquid chromatography, was then directly microsequenced, and ³H-incorporation associated with each individual amino acid was analysed by scintillation counting. The results confirmed selective transfer of methyl-label to Lys 9 (Fig. 4c), showing that Suv39h1 is a highly site-specific HMTase for the H3 N terminus *in vitro*.

Lys 9 methylation and Ser 10 phosphorylation

Position-specific modifications of the H3 N terminus have been correlated with distinct chromatin-based outputs, such as transcriptional regulation (Lys 14 acetylation), histone deposition (Lys 9 acetylation) and chromosome condensation/segregation (Ser 10 phosphorylation)¹. We analysed whether methylation of Lys 9 would possibly interfere with this ‘histone code’ and whether it might be influenced by pre-existing modifications at these sites. We performed *in vitro* HMTase assays with GST–SUV1(82–412) and H3 N-terminal peptides that were either unmodified, phosphorylated at Ser 10 (S10-phos), di-methylated at Lys 9 (K9-dimeth), or acetylated at Lys 9 (K9-Ac) or Lys 14 (K14-Ac). We visualized the processed reaction products by fluorography and quantified them with a Phosphorimager, setting the amount of SUV39H1-dependent methylation of the unmodified peptide as 100% (Fig. 5a). As expected, a Lys 9 dimethylated position severely compromised enzymatic activity, and a Lys 9 trimethylated peptide was no longer a substrate (data not shown). Similarly, pre-existing acetylation at Lys 9 prevented SUV39H1-dependent methylation, whereas acetylation at Lys 14 did not significantly interfere. Unexpectedly, the S10-phos peptide could also not be methylated, suggesting that Ser 10 phosphorylation inhibits SUV39H1-dependent methylation of the adjacent Lys 9.

A mitotic kinase (Ipl1/aurora) that selectively phosphorylates Ser 10 in H3 has recently been identified⁹. We used recombinant Ipl1/aurora kinase to examine a potential reciprocal interference of Ser 10 phosphorylation by pre-existing methylation at Lys 9 with our set of modified H3 N-terminal peptides. These *in vitro* kinase assays indicated that the K14-Ac peptide represents an even better substrate than the unmodified N terminus, whereas a Ser 10 phosphorylated position allowed only residual activity. Notably, di-methylation at Lys 9, but not acetylation at Lys 9, significantly reduced the substrate quality of the H3 N terminus, and 35% or less of Ipl1/kinase activity was detected with the K9-dimeth peptide (Fig. 5b). Together with the above enzymatic analyses of SUV39H1 HMTase preference, these data provide strong experimental evidence for an interdependence of site-specific histone tail modifications.

Modification of H3 in native chromatin

Murine *Suv39h* genes are encoded by two loci (*Suv39h1* and *Suv39h2*; D.O’C. *et al.*, manuscript in preparation), both of which have recently been disrupted in the mouse germ line. Although single *Suv39h1* and *Suv39h2* null mice are viable, double *Suv39h* deficient mice are born at only about 20–25% of the expected mendelian ratios and are growth retarded, showing that *Suv39h* has an important function during mouse development (D.O’C. *et al.*, manuscript in preparation). To begin to uncover the role(s) of Suv39h-dependent H3 methylation *in vivo*, we generated primary mouse embryonic fibroblasts (PMEFs) from *Suv39h* double null fetuses. Cell-cycle profiles of wild-type and *Suv39h* double null PMEFs indicated that similar percentages of cells are in S and G2/M phases (Fig. 6a), but that *Suv39h* double null PMEFs have a reduced G1 index and an increased proportion of cells with aberrant nuclear morphologies, reminiscent of division defects during mitosis. For example, *Suv39h* double null PMEFs contain twofold or more elevated numbers of cells with micro- and polynuclei, and are further characterized by cell subpopulations with oversized nuclei or a weak definition of heterochromatin that appears in only a few unusually condensed foci (Fig. 6b).

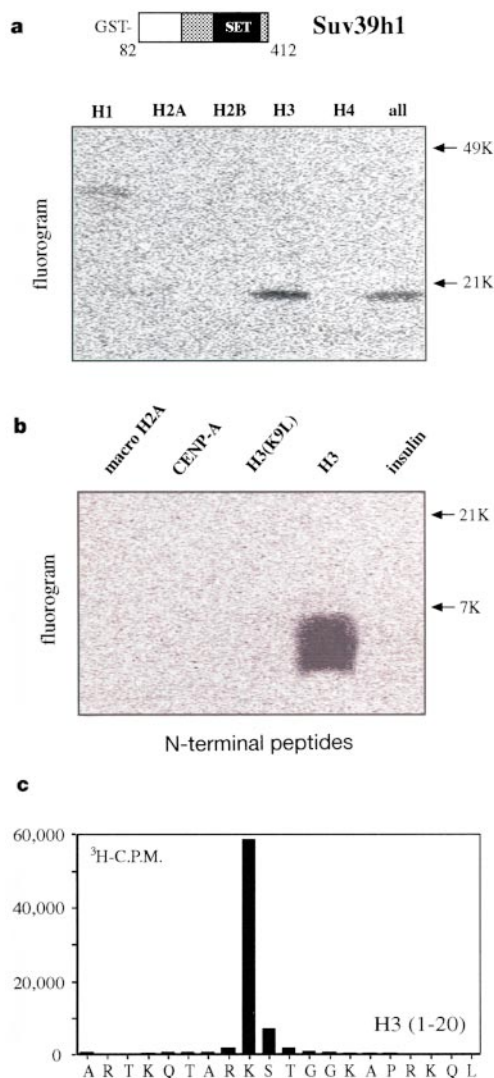


Figure 4 Lysine 9 of the H3 N terminus is the principal site for *in vitro* methylation by recombinant Suv39h1. **a**, About 10 µg of GST–Suv1(82–412) (murine Suv39h1) were used in *in vitro* HMTase reactions with individual histones. Only the fluorogram is shown. **b**, *In vitro* methylation assays using GST–Suv1(82–412) as enzyme and the indicated N-terminal peptides of wild-type H3, mutated H3(K9L), CENP-A, macroH2A or insulin as substrates. **c**, Automated sequencing of the wild-type H3 N-terminal peptide (residues 1–20) that had been methylated *in vitro* by recombinant GST–Suv1(82–412). The tritium incorporation of individual amino acids identified at each successive round of microsequencing is shown.

Phosphorylation at Ser 10 in the N-terminal tail of H3 (phosH3) is required for condensation and subsequent segregation of chromosomes⁵. During the cell cycle, phosH3 initiates within pericentric heterochromatin in late G2 and then progresses along

the entire chromosomes during mitosis⁴⁰. In wild-type PMEFs, about 7% of cells stain positive for the characteristic, heterochromatin-associated phosH3 foci, as detected by indirect immunofluorescence with anti-phosH3 antibodies (Fig. 6c, right panel). In

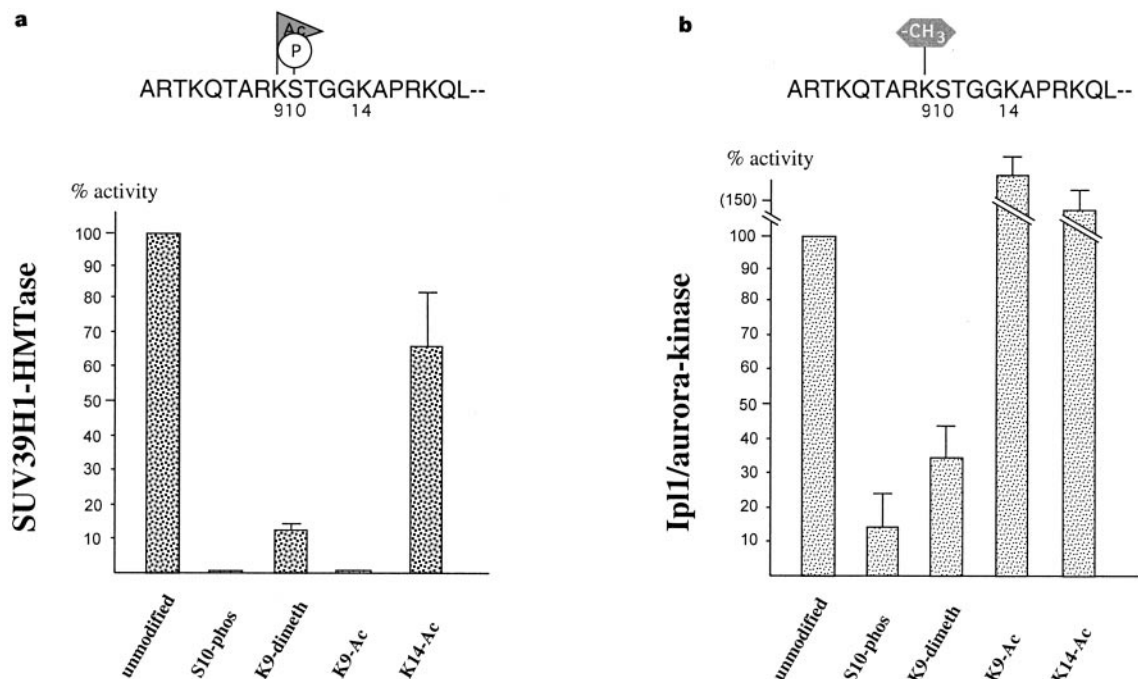


Figure 5 Interdependence of Lys 9 methylation and Ser 10 phosphorylation in the H3 N terminus. **a**, *In vitro* HMTase assays with GST-SUV1(82–412) and H3 N-terminal peptides that are either unmodified or phosphorylated (S10-phos), di-methylated (K9-dimeth) or acetylated (K9-Ac or K14-Ac) at the indicated positions. **b**, *In vitro* kinase

assays with recombinant His₁₀-Ipl1/aurora and the same set of H3 N-terminal peptides. For **a** and **b**, the percentage of enzyme activity detected with each individual peptide is plotted, setting the reaction with unmodified H3 peptide to 100%. For both enzymes, inhibiting modifications in the H3 N terminus are shown on top of the respective panels.

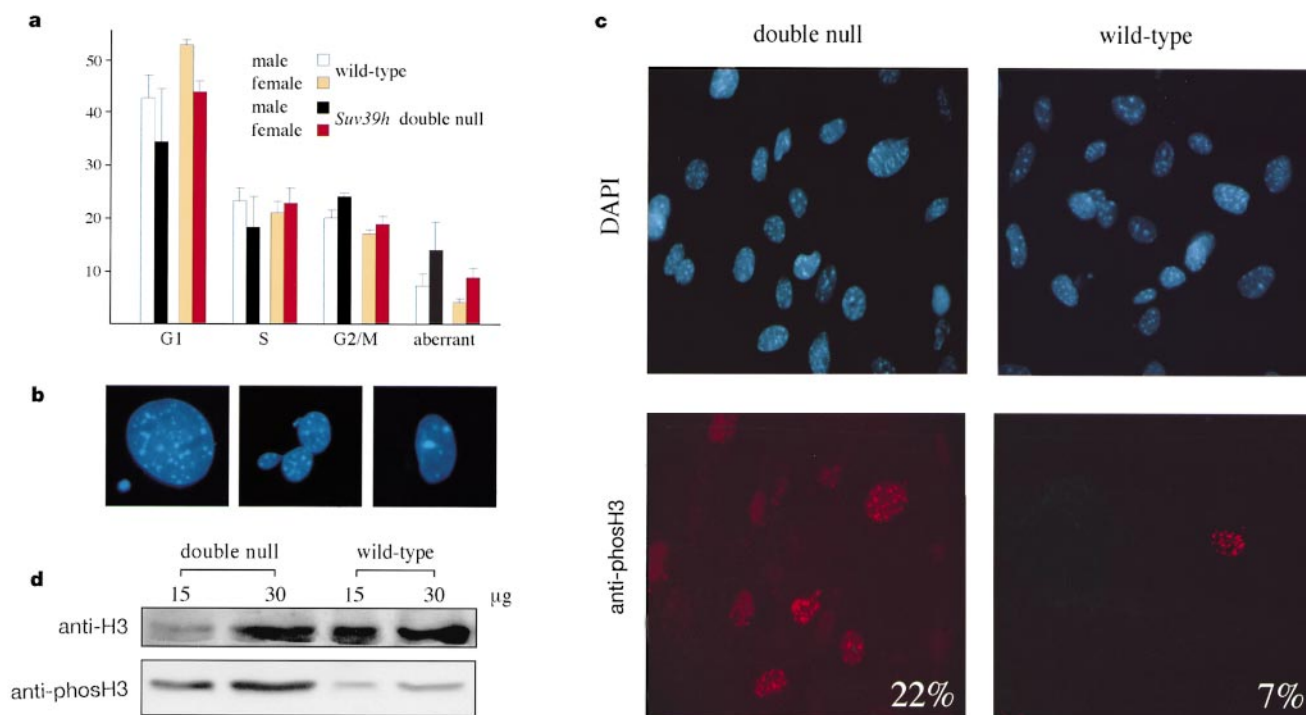


Figure 6 *Suv39h*-dependent modification of the H3 N terminus in native chromatin. **a**, Percentages of cell-cycle stages of passage 3 wild-type and *Suv39h* double null primary mouse embryonic fibroblasts (PMEFs). **b**, Representative images of aberrant mitoses in *Suv39h* double null PMEFs detected by β -tubulin (not shown) and DAPI staining (blue). Also shown (right image) is a nucleus exemplifying the unusual definition of heterochromatin in a subpopulation of *Suv39h* double null PMEFs. Original magnification

of all images, $\times 630$. **c**, Interphase chromatin staining with anti-phosH3 antibodies and CY3-conjugated secondary antibodies (red). DNA was counterstained with DAPI (blue). Roughly 1,000 cells were counted to evaluate the percentage (shown bottom right) of anti-phosH3 positive cells. **d**, Quantitative western analysis with 15 μ g and 30 μ g of total nuclear proteins immuno-blotted with anti-H3 and anti-phosH3 antibodies.

contrast, in *Suv39h* double null PMEFs about 22% of cells contain phosH3-positive foci (Fig. 6c, left panel), although their definition appears in many small speckles that do not always overlap with DAPI-dense material (data not shown). This result suggested that the overall levels of phosH3 may be enhanced in *Suv39h* double null PMEFs. We therefore determined the relative abundance of phosH3 in precalibrated nuclear extracts with anti-phosH3-specific antibodies. This quantification indicated a significantly higher level of phosH3 in *Suv39h* double null cells than in wild-type controls (Fig. 6d). These *in vivo* data complement our enzymatic *in vitro* analyses, and confirm a model in which methylation of Lys 9 in H3 impairs phosphorylation of adjacent Ser 10 at specific regions in native chromatin.

HMTase versus MTase

We have shown that SU(VAR)3-9-related proteins possess a selective H3 HMTase activity towards free histones *in vitro*. This H3 HMTase activity, however, was not detected with other constructs containing SET domains, for example EZH2 or HRX. Although both EZH2 and HRX contain clustered cysteines in their N-terminal halves^{30,31,36,37}, they lack either the C-terminal tail (EZH2) or the SET domain-associated cysteine-rich region (HRX). As observed for HATs, it is likely that the targets for SET-domain-dependent methyltransferase (MTase) activity are probably not restricted to histones. For example, the functionally characterized *Pisum sativum* lysine MTase methylates the large subunit of a metabolic enzyme^{33,34}, and an *S. cerevisiae* gene encoding a cytochrome *c* MTase activity⁴¹ also contains a SET domain (see Methods). We therefore propose a generic MTase quality for the SET domain, whose substrate specificity would largely depend upon combinations with additional protein modules. Alternatively, some SET domains may intrinsically lack MTase activity, similar to described anti-phosphatases that can counteract the roles of the catalytically active enzyme(s)⁴². This interpretation is particularly intriguing, given that an anti-phosphatase has been shown to be a SET domain binding factor^{42,43}, and that the SET domain is present in chromatin regulators of apparently antagonistic functions^{19,27,30,31}.

A 'histone code' for higher-order chromatin

SUV39H1 and Suv39h1 are modular proteins that are enriched at heterochromatin^{24,35}, thereby directing their highly selective H3 HMTase activity to localized chromatin regions. Indeed, Lys 9 methylation appears nearly unchanged in bulk H3 preparations from nuclear extracts of wild-type and *Suv39h* double null PMEFs (data not shown), indicating that Lys 9 is more widely methylated by additional HMTases. Unexpectedly, methylation of Lys 9 interferes with Ipl1/aurora-dependent phosphorylation of Ser 10, but is also inhibited by pre-existing modifications in the H3 N terminus. These data argue that the *in vivo* substrate specificity for Suv39h-dependent methylation would be maximal for an unmodified H3 N terminus, which is primarily presented after DNA replication and nucleosome assembly. Notably, chromatin assembly factor 1 interacts with murine HP1 proteins⁴⁴, which have previously been shown to complex with SUV39H1 or Suv39h1 (ref. 24).

In addition to modulating phosH3 distribution, overexpression of SUV39H1 also induces ectopic heterochromatin³⁵, whereas disruption of *Suv39h* function prevents heterochromatin association of HP1 proteins (data not shown), suggesting that Lys 9 methylation in H3 might impart a preferred binding surface for heterochromatin-specific proteins. This 'heterochromatic' affinity would be highest in the absence of additional modifications and may even be self-propagating within transcriptionally less-permissive chromatin regions, consistent with a maximal H3 HMTase activity of SUV39H1 towards an unmodified H3 N terminus. Our data predict that a regional increase in the methylation profile of Lys 9 in H3, probably in conjunction with a reduced acetylation status of the H4 N terminus¹⁻³, represents a key determinant in defining a 'histone

code' for heterochromatin. On the basis of our results, we propose that SU(VAR)3-9 related proteins provide important enzymatic activities towards the induction and assembly of higher-order chromatin. □

METHODS

Sequence alignments and secondary-structure predictions

The SET domains of human SUV39H1 (ref. 24), *Drosophila* SU(VAR)3-9 (ref. 19) and *S. pombe* CLR4 (ref. 22) were used as a multiple-starting alignment for database similarity searches using Profile⁴⁵, hidden Markov⁴⁶ and position-specific iterative BLAST⁴⁷ methods (representative listings are available from the SET domain page of the SMART WWW server³²). These searches revealed significant similarities to six plant proteins (GenBank accession numbers Q43088, O65218, P94026, O80013, AAC29137 and AC007576_12) described as putative lysine N-methyltransferases. For example, a PSI-BLAST search using the *S. pombe* hypothetical protein SPAC3c7.09 as query identified these plant sequences and well known SET domain sequences within ten rounds using an E-value inclusion threshold of 0.001. The same search also revealed the presence of a SET domain in YHR109w (which is known to encode a cytochrome *c* MTase⁴¹) within three rounds. Consensus secondary structures were predicted by described algorithms⁴⁸.

Epitope-tagged SUV39H1 proteins in HeLa cells

The HeLa cell lines overexpressing full-length (myc)₃-SUV39H1 (3–412) or (myc)₃-Nchromo (3–118) have been described³⁵. Nuclear extracts were immunoprecipitated with anti-myc antibody beads²⁴, and about 1–3 µg of matrix-bound (myc)₃-tagged SUV39H1 proteins were used for *in vitro* HMTase assays.

Generation and purification of GST fusion proteins

The GST-Suv1(82–412) product expressed from the pGEX-2T vector (Pharmacia) has been described²⁴. Additional GST constructs were generated by transferring *Bam*HI–*Eco*RI PCR amplicons into pGEX-2T, encoding in-frame fusions for Suv1(7–221), SUV1(82–412), SUV1(82–378)ΔC-tail, SUV1(255–412)Δcys, Suv2(157–477) (D. O'C. *et al.*, manuscript in preparation), CLR4(127–490)²², EZH2(382–747)³⁶ and HRX(3643–3969)³⁷. Short internal deletions (ΔNHSCDPN_{323–329} and ΔGEELTFDY_{358–365}) or point mutations (see Fig. 2c) within the 320HΦΦNHSC₃₂₆ motif were directly engineered in the GST-SUV1(82–412) plasmid by double PCR mutagenesis. All constructs were confirmed by sequencing.

Recombinant proteins were expressed in 11 cultures of *E. coli* strain BL21 and solubilized in 10 ml RIPA buffer (20 mM Tris pH 7.5, 500 mM NaCl, 5 mM EDTA, 1% NP-40 and 0.5% sodium deoxycholate containing a full set of protease inhibitors (Boehringer Mannheim) and lysozyme (5 mg ml⁻¹; Sigma)) by freeze-thawing in liquid N₂, followed by sonication. Soluble proteins were cleared by centrifugation, purified with 800 µl glutathione Sepharose beads (Pharmacia) and washed twice in RIPA buffer. Protein concentration was determined by Coomassie staining of SDS-PAGE gels. Matrix-bound fusion proteins were used immediately for *in vitro* HMTase assays or stored at 4 °C.

In vitro histone methyltransferase assay

In vitro HMTase reactions were modified on the basis of described protocols⁷ and carried out in a volume of 50 µl of methylase activity buffer (MAB: 50 mM Tris pH 8.5, 20 mM KCl, 10 mM MgCl₂, 10 mM β-mercaptoethanol, 250 mM sucrose), containing 10 µg of free histones (mixture of H1, H3, H2B, H2A and H4; Boehringer Mannheim) as substrates and 300 nCi S-adenosyl-[methyl-¹⁴C]-L-methionine (25 µCi ml⁻¹) (Amersham) as methyl donor. We routinely used 10 µg of matrix-bound GST fusion proteins to assay for HMTase activity. After incubation for 60 min at 37 °C, reactions were stopped by boiling in SDS loading buffer, and proteins were separated by 15% or 18% SDS-PAGE and visualized by Coomassie staining and fluorography.

HMTase assays with individual histones (Boehringer Mannheim), insulin (Sigma) or N-terminal peptides were performed with 5 µg of substrate. We used the following peptides: wild-type N terminus of human histone H3 (ARTKQTARKSTGGKAPRKQL) and mutant peptide which changes Lys 9 (underlined) to leucine; N terminus of human CENP-A (MGPRRRSRKPEAPRRRSPP)³⁸; N terminus of rat macro-H2A (MSSRGKKKSTKTSRSKAG)³⁹. Modified H3(1–20) peptides (see Fig. 5) were synthesized using phosphorylated serine, di- or tri-methylated lysine and acetylated lysine (all Bachem or NovaBiochem).

Peptide microsequencing of the *in vitro* methylated wild-type H3 N-terminal peptide and determination of ³H-incorporation of individual amino acids by scintillation counting was done as described⁷.

In vitro Ipl1/aurora kinase assay

Purification of recombinant His₁₀-Ipl1 and analysis of kinase assays with H3 N-terminal peptides by scintillation counting was done as described⁹.

Generation and analysis of *Suv39h* double null PMEFs

Primary mouse embryonic fibroblasts were derived from day E12.5 *Suv39h* double null fetuses obtained after intercrossing *Suv39h1*^{-/-}/*Suv39h2*^{-/-} compound mutant mice (D.O'C. *et al.*, manuscript in preparation). As controls, PMEFs were prepared from wild-type fetuses of the same genetic background. For cell-cycle profile analysis by FACS, passage three PMEFs were analysed as described⁴⁹. Staining of unextracted PMEF

interphase chromatin with anti-phosH3 (ref. 40) antibodies and counterstaining of DNA with 4',6'-diamidino-2-phenylindole (DAPI; Sigma) was done as described³⁵. For the biochemical analysis, total nuclear extracts were precalibrated by Ponceau staining, immuno-blotted with anti-H3 (Upstate Biotechnology) and anti-phosH3 (ref. 40) antibodies and visualized by peroxidase staining using enhanced chemiluminescence (Amersham).

Received 24 February; accepted 9 June 2000.

- Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
- Grunstein, M. Yeast heterochromatin: regulation of its assembly and inheritance by histones. *Cell* **93**, 325–328 (1998).
- Turner, B. M. Histone acetylation as an epigenetic determinant of long-term transcriptional competence. *Cell. Mol. Life Sci.* **54**, 21–31 (1998).
- Karpen, G. & Allshire, R. C. The case for epigenetic effects on centromere identity and function. *Trends Genet.* **13**, 489–496 (1997).
- Wei, Y., Lanlan, Y., Bowen, J., Gorovsky, M. A. & Allis, C. D. Phosphorylation of histone H3 is required for proper chromosome condensation and segregation. *Cell* **97**, 99–109 (1999).
- Chen, D. *et al.* Regulation of transcription by a protein methyltransferase. *Science* **284**, 2174–2177 (1999).
- Strahl, B. D., Ohba, R., Cook, R. G. & Allis, C. D. Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in *Tetrahymena*. *Proc. Natl Acad. Sci. USA* **96**, 14967–14972 (1999).
- Sassone-Corsi, P. *et al.* Requirement of Rsk-2 for epidermal growth factor-activated phosphorylation of histone H3. *Science* **285**, 886–891 (1999).
- Hsu, J.-Y. *et al.* Mitotic phosphorylation of histone H3 is governed by Ipl1/aurora kinase and Glc7/PP1 phosphatase in budding yeast and nematodes. *Cell* **102**, 1–20 (2000).
- Reuter, G. & Spierer, P. Position-effect variegation and chromatin proteins. *BioEssays* **14**, 605–612 (1992).
- Allshire, R. C., Nimmo, E. R., Ekwall, K., Javerzat, J. P. & Cranston, G. Mutations derepressing silent centromeric domains in fission yeast disrupt chromosome segregation. *Genes Dev.* **9**, 218–233 (1995).
- Wallrath, L. L. Unfolding the mysteries of heterochromatin. *Curr. Opin. Genet. Dev.* **8**, 147–153 (1998).
- De Rubertis, F. *et al.* The histone deacetylase RPD3 counteracts genomic silencing in *Drosophila* and yeast. *Nature* **384**, 589–591 (1996).
- Grewal, S. I., Bonaduce, M. J. & Klar, A. J. Histone deacetylase homologs regulate epigenetic inheritance of transcriptional silencing and chromosome segregation in fission yeast. *Genetics* **150**, 563–576 (1999).
- Baksa, K. *et al.* Mutations in the protein phosphatase 1 gene at 87B can differentially affect suppression of position effect variegation and mitosis in *Drosophila melanogaster*. *Genetics* **135**, 117–125 (1993).
- Larsson, J., Zhang, J. & Rasmuson-Lestander, Å. Mutations in the *Drosophila melanogaster* gene encoding S-adenosyl methionine suppress position-effect variegation. *Genetics* **143**, 887–896 (1996).
- Eissenberg, J. C., Morris, G. D., Reuter, G. & Hartnett, T. The heterochromatin-associated protein HP-1 is an essential protein in *Drosophila* with dosage-dependent effects on position-effect variegation. *Genetics* **131**, 345–352 (1992).
- Cléard, F., Delattre, M. & Spierer, P. SU(VAR)3-7, a *Drosophila* heterochromatin-associated protein and companion of HP1 in the genomic silencing of position-effect variegation. *EMBO J.* **16**, 5280–5288 (1997).
- Tschiersch, B. *et al.* The protein encoded by the *Drosophila* position-effect variegation suppressor gene *Su(var)3-9* combines domains of antagonistic regulators of homeotic gene complexes. *EMBO J.* **13**, 3822–3831 (1994).
- Schotta, G. & Reuter, G. Controlled expression of tagged proteins in *Drosophila* using a new modular P-element vector system. *Mol. Gen. Genet.* **262**, 916–920 (2000).
- Henikoff, S. in *Epigenetic Mechanisms of Gene Regulation* (eds Russo, V. A. E., Martienssen, R. A. & Riggs, A. D.) 319–334 (CSHL, New York, 1996).
- Ivanova, A. V., Bonaduce, M. J., Ivanov, S. V. & Klar, A. J. S. The chromo and SET domains of the Clr4 protein are essential for silencing in fission yeast. *Nature Genet.* **19**, 192–195 (1998).
- Ekwall, K. *et al.* Mutations in the fission yeast silencing factors *clr4⁺* and *riki⁺* disrupt the localisation of the chromo domain protein Swi6p and impair centromere function. *J. Cell Sci.* **109**, 2637–2648 (1996).
- Aagaard, L. *et al.* Functional mammalian homologues of the *Drosophila* PEV modifier *Su(var)3-9* encode centromere-associated proteins which complex with the heterochromatin component M31. *EMBO J.* **18**, 1923–1938 (1999).
- Aasland, R. & Stewart, A. F. The chromo shadow domain, a second chromodomain in heterochromatin-binding protein1, HP1. *Nucleic Acids Res.* **23**, 3168–3173 (1995).
- Koonin, E. V., Zhou, S. & Lucchesi, J. C. The chromo superfamily: new members, duplication of the chromo domain and possible role in delivering transcription regulators to chromatin. *Nucleic Acids Res.* **23**, 4229–4232 (1995).
- Jenuwein, T., Laible, G., Dorn, R. & Reuter, G. SET-domain proteins modulate chromatin domains in eu- and heterochromatin. *Cell. Mol. Life Sci.* **54**, 80–93 (1998).
- Ball, L. J. *et al.* Structure of the chromatin binding (chromo) domain from mouse modifier protein 1. *EMBO J.* **16**, 2473–2481 (1997).
- Platero, J. S., Harnett, T. & Eissenberg, J. C. Functional analysis of the chromo domain of *HP-1*. *EMBO J.* **14**, 3977–3986 (1995).
- Jones, R. S. & Gelbart, W. M. The *Drosophila* Polycomb-group gene *Enhancer of zeste* contains a region with sequence similarity to *trithorax*. *Mol. Cell. Biol.* **13**, 6357–6366 (1993).
- Stassen, M. J., Bailey, D., Nelson, S., Chinwalla, V. & Harte, P. J. The *Drosophila trithorax* protein contains a novel variant of the nuclear receptor type DNA binding domain and an ancient conserved motif found in other chromosomal proteins. *Mech. Dev.* **52**, 209–223 (1995).
- Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P. & Bork, P. SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* **28**, 231–234 (2000).
- Klein, R. R. & Houtz, R. L. Cloning and developmental expression of pea ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit N-methyltransferase. *Plant Mol. Biol.* **27**, 249–261 (1995).
- Zheng, Q., Simel, E., Klein, P., Royer, M. & Houtz, R. L. Expression, purification, and characterization of recombinant ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit N⁶-methyltransferase. *Protein Exp. Purif.* **14**, 104–112 (1998).
- Melcher, M. *et al.* Structure–function analysis of SUV39H1 reveals a dominant role in heterochromatin organization, chromosome segregation and mitotic progression. *Mol. Cell. Biol.* **20**, 3728–3741 (2000).
- Laible, G., *et al.* Mammalian homologues of the Polycomb-group gene *Enhancer of zeste* mediate gene silencing in *Drosophila* heterochromatin and at *S. cerevisiae* telomeres. *EMBO J.* **16**, 3219–3232 (1997).
- Tkachuk, D. C., Kohler, S. & Cleary, M. L. Involvement of a homolog of *Drosophila trithorax* by 11q23 chromosomal translocations in acute leukemias. *Cell* **71**, 691–700 (1992).
- Sullivan, K. F., Hechenberger, M. & Masri, K. Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. *J. Cell Biol.* **127**, 581–592 (1994).
- Pehrson, J. R. & Fried, V. A. MacroH2A, a core histone containing a large nonhistone region. *Science* **257**, 1398–1400 (1992).
- Hendzel, M. J. *et al.* Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. *Chromosoma* **106**, 348–360 (1997).
- Martzen, M. R. *et al.* A biochemical genomics approach for identifying genes by the activity of their products. *Science* **286**, 1153–1155 (1999).
- Cui, X. *et al.* Association of SET domain and myotubularin-related proteins modulates growth control. *Nature Genet.* **18**, 331–337 (1998).
- Firestein, R., Cui, X., Huie, P. & Cleary, M. L. SET domain-dependent regulation of transcriptional silencing and growth control by SUV39H1, a mammalian ortholog of *Drosophila* Su(var)3-9. *Mol. Cell. Biol.* **20**, 4900–4909 (2000).
- Murzina, N., Verrault, A., Laue, E. & Stillman, B. Heterochromatin dynamics in mouse cells: interaction between chromatin assembly factor1 and HP1 proteins. *Mol. Cell* **4**, 529–540 (1999).
- Birney, E., Thompson, J. D. & Gibson, T. J. PairWise and SearchWise: finding the optimal alignment in a simultaneous comparison of a protein profile against all DNA translation frames. *Nucleic Acids Res.* **24**, 2730–2739 (1996).
- Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14**, 755–763 (1998).
- Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Frishman, D. & Argos, P. Seventy-five percent accuracy in protein secondary structure prediction. *Proteins Struct. Funct. Genet.* **27**, 329–335 (1997).
- Xu, X. *et al.* Centrosome amplification and a defective G₂-M cell cycle checkpoint induce genetic instability in BRCA1 exon 11 isoform-deficient cells. *Mol. Cell* **3**, 389–395 (1999).

Acknowledgements

We would like to thank M. Cleary for providing a partial *HRX* cDNA, R. Allshire for the *Clr4* cDNA, I. Gorny for peptide synthesis and R. G. Cook for automated sequencing of the H3 N-terminal peptide. During the course of this work, E. V. Koonin and L. Aravind have independently discovered homology of the SET domain with plant methyltransferases. We acknowledge M. Doyle for the contribution to the *Suv39HZ* knockout and thank K. Nasmyth and M. Busslinger for helpful comments and critical reading of the manuscript. Research in the laboratory of C.D.A. is funded by grants from the NIH to C.D.A. and B.D.S. Z.-W.S. is supported by a postdoctoral cancer training grant from the University of Virginia Cancer center. Research in the laboratory of T.J. is supported by the IMP, the Austrian Research Promotion Fund and the Vienna Business Agency.

Correspondence and requests for materials should be addressed to T.J. (e-mail: jenuwein@nt.imp.univie.ac.at).

Evaporation in the young solar nebula as the origin of 'just-right' melting of chondrules

Bosmat A. Cohen, Roger H. Hewins & Yang Yu

Department of Geological Sciences, Rutgers University, Piscataway, New Jersey 08855, USA

Chondrules^{1–5} are millimetre-sized, solidified melt spherules formed in the solar nebula by an early widespread heating event of uncertain nature^{6–8}. They were accreted into chondritic asteroids, which formed about 4.56 billion years ago and have not experienced melting or differentiation since that time. Chondrules have diverse chemical compositions, corresponding to liquidus temperatures^{1,4,9} in the range 1,350–1,800 °C. Most chondrules, however, show porphyritic textures (consisting of large crystals in a distinctly finer grained or glassy matrix), indicative of melting within the narrow range 0–50 °C below the liquidus^{9,10}. This suggests an unusual heating mechanism for chondrule precursors, which would raise each individual chondrule to just the right temperature (particular to individual bulk composition) in order to form porphyritic textures. Here we report the results of isothermal melting of a chondritic composition at nebular pressures. Our results suggest that evaporation stabilizes porphyritic textures over a wider range of temperatures below the liquidus (about 200 °C) than previously believed, thus removing the need for individual chondrule temperature buffering. In addition, we show that evaporation explains many chondrule bulk and mineral compositions that have hitherto been difficult to understand.

Chondrules exhibit a wide range of chemical compositions. Some are enriched in magnesium and depleted in moderately volatile

elements (type I) and some are enriched in iron and in moderately volatile elements (type II)¹. Furthermore, silica content in chondrules varies as well between silica-rich chondrules (suffix B) and silica-poor ones (suffix A)^{4,5}. Porphyritic textures are common in at least 80% of all chondrules¹¹. These textures are formed by cooling of melts heated close to their liquidus (the temperature at which all solid phases disappear)^{9,10}. Temperatures in this range (0–50 °C below the liquidus) destroy nearly all of the solid mineral phases that can serve as nucleation sites. Upon cooling, the few remaining nuclei grow large well-formed crystals characteristic of porphyritic textures. If melts are superheated above their liquidus, all nuclei are destroyed and barred textures or glasses are formed upon cooling. Conversely, heating to temperatures lower than 50 °C below the liquidus will lead to the preservation of many nuclei and the resulting texture upon cooling would be granular or cryptoporphyratic¹².

If chondrule bulk compositions do not change during melting, the peak temperature must correlate with composition to produce porphyritic texture. However, each of the many proposed melting mechanisms^{2,6–8}, including shock waves, lightning discharges and flares, would heat all precursors to approximately the same temperature. No known heating mechanism has the ability to buffer each chondrule to the exact temperature needed for the formation of porphyritic textures. Chondrules possibly show chemical variations as a result of evaporation at nebular pressures^{3,12–14}. We have therefore tested the effect on composition and texture of heating at constant temperature (close to but below the liquidus of their precursors) for different lengths of time.

In our experiments, we evaporated an anhydrous mineral mixture matching average Solar System dust (CI—carbonaceous chondrite Ivona type)^{15,16} for different lengths of time (1 to 18 h) at 1,580 °C in a vacuum furnace¹⁷. As it is believed that chondrules were formed under nebular conditions, that is, low pressures and low oxygen fugacity, the experiments were performed at a p_{H_2} of 1.3×10^{-4} atm. Charges were quenched at the end of each experiment.

Our charges for 1–4 h of evaporation resemble type II chondrules

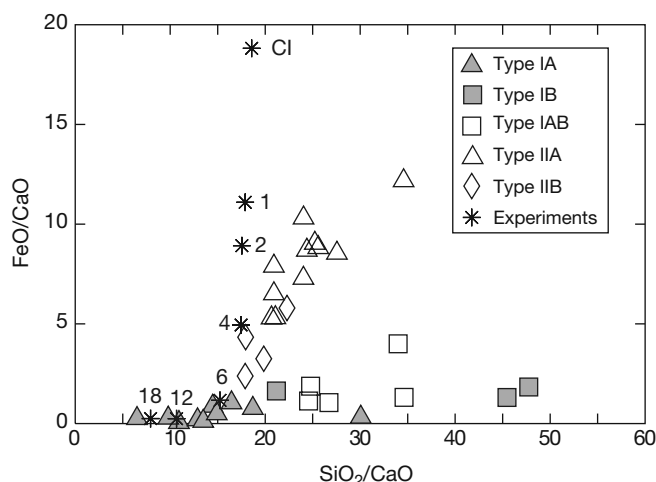


Figure 1 Evaporative loss of Fe and Si in experimental residues produces compositions similar to those of natural chondrules⁴. Starting material (marked CI) is composed of 31.6% SiO₂, 22.1% MgO, 20.2% FeS, 16.8% FeO, 3.7% C, 2.02% Al₂O₃, 1.9% CaO, 1.05% Na₂O and 0.03% K₂O and its calculated liquidus is 1,630 °C. It was prepared by mixing the minerals olivine, diopside, plagioclase, troilite, plus Sarro-Lorraine kerogen. Pulverized starting material, pressed into 2-mm-wide pellets, were hung on an iron-plated rhenium wire and inserted into a vertical Deltech furnace. Experiments were conducted at 1,580 °C and at a p_{H_2} of 1.3×10^{-4} atm. Numbers near asterisks indicate heating duration in hours. SiO₂/CaO and FeO/CaO mass ratios of the experimental residues of 6, 12 and 18 h runs are similar to type IA and 4-hour-run residues resemble type IIB.

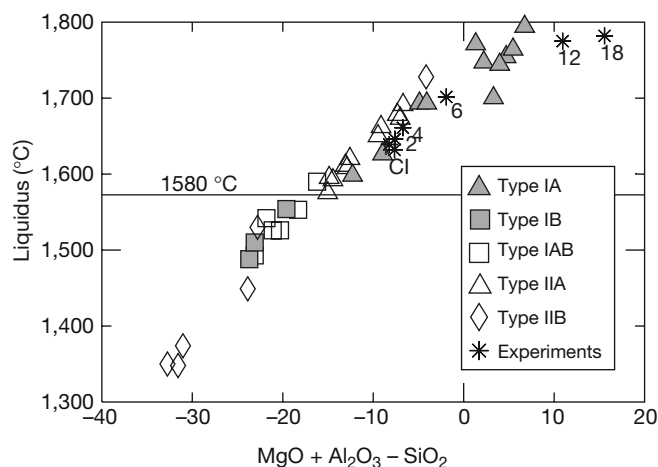


Figure 2 Experimental residues formed at 1,580 °C and natural chondrules¹³ show a similar relationship between calculated liquidus temperatures and bulk composition. Experiment durations are marked near asterisks. As evaporation time increases the liquidus temperature of the residue increases as well, following the natural chondrule trend. Producing the chondrules on the left of CI would require more pyroxene-rich precursors.

(Fig. 1). After 6 h of evaporation the charges lost roughly 85% of the FeO and the evaporation of SiO₂ became dominant, giving type I compositions. As Fe evaporates the Mg content in olivine crystals (69 mole per cent Mg₂SiO₄ in the starting material) rises, giving almost pure forsterite after 12 h. As olivine crystals become more Mg-rich their CaO content increases as well, due to the relative increase of CaO in the melt. The same trend is found in olivine grains in natural chondrules⁴. However, the origin of the very Ca-rich forsterite grains has also been explained by condensation¹⁸.

By comparing the composition of our starting material with that of the residual charges from our experiments we can estimate the mass fraction lost during evaporation. The total mass loss for the first 6 h of evaporation considering total loss of FeS (20%), Na₂O (1%) and FeO (17%) is 38%. Assuming no Mg is lost, about 40% of the SiO₂ (~13% of the total mass) is lost, raising the total mass loss

to about 51% for the longest run. This is a minimum, as after 12 h a little Mg is lost. These results agree with calculations³ showing that type I chondrules are about 50% smaller in size relative to type II.

Our evaporation residues are similar to type IIA and IA chondrules in composition (Fig. 1), and also in calculated liquidus temperature (Fig. 2). Fe-poor, Si-rich precursors are needed for type IAB (AB indicates similar proportions of olivine and pyroxene) and type IB chondrules, if they are to be explained by evaporation. Such chondrule compositions that fall off the evaporation trend may result from recycling of Si-rich chondrule debris¹⁹. More generally, any precursor heterogeneity would explain the composition spread and, though we used an average solar composition, we do not believe that nebular condensates were perfectly homogeneous. Variations in the ratio between olivine to pyroxene in precursors would explain the compositions of such chondrules.

A timescale of hours for chondrule formation^{10,20–22} is consistent with evaporation to produce chondrules, with shorter times for type IIA chondrules and longer for type IA. Higher cooling rates would be expected for type II relative to type I chondrules. Cooling rates have been estimated for type IIA chondrules based on Fe–Mg zoning in olivine grains and are relatively high, at 10–1,000 °C h^{–1} (ref. 10). This method fails for type I chondrules because they are FeO-poor and their olivine grains are not zoned. However, pyroxene exsolution properties in some type I chondrules were used for calculating cooling rates and gave about 10 °C h^{–1} or less²¹. A correlation between cooling rate and composition is therefore possible, though it has not yet been demonstrated conclusively.

All our evaporative residues are porphyritic. Experiments performed at 1 atm, with no evaporation, show that porphyritic textures require a relatively narrow temperature range, within about 50 °C of the liquidus (Fig. 3a). Our experiments are performed at low pressures; during heating to about 50 °C below the liquidus (of the starting material), olivine grains dissolve leaving many small relicts, which Ostwald-ripen²³ to coarse porphyritic texture in less than 1 h. Olivine abundance decreases at first as FeO is lost and increases later as SiO₂ is lost, as expected³. However, as the number of crystals in the melt remains roughly constant as evaporation proceeds, the porphyritic texture is hardly modified and liquidus temperatures can climb well above the actual temperature (Fig. 3b). Consequently porphyritic textures formed at temperatures from about 50 °C below to more than 200 °C below the (final) liquidus temperature of the residual spherule. Thus, porphyritic chondrules formed by open-system melting do not need 'just right' heating but rather melting temperatures close to the liquidus of CI-like precursors.

If porphyritic textures in chondrules were to form by open-system evaporation of CI starting material, then according to Fig. 2, chondrules with liquidus temperatures below about 1,500 °C (that is, Si-rich compositions), are less likely to exhibit porphyritic textures. This result perfectly matches the observed abundance in natural chondrules, which shows that the majority of chondrules with liquidus temperatures below about 1,500 °C have excentro-radial, barred and cryptocrystalline textures^{1,9}. The relatively limited number of porphyritic textures in Si-rich chondrules could have formed by dust seeding of total melts during cooling. Producing porphyritic texture by introducing dust as nucleation sites has been experimentally demonstrated²⁴. □

Received 13 March; accepted 7 June 2000.

- McSweeney, H. Y. Jr Chemical and petrographic constraints on the origin of chondrules and inclusions in carbonaceous chondrites. *Geochim. Cosmochim. Acta* **41**, 1843–1860 (1977).
- Grossman, J. N. in *Meteorites and the Early Solar System* (eds Kerridge, J. F. & Matthews, M. S.) 680–696 (Univ. Arizona Press, Tucson, 1988).
- Sears, D. W. G., Huang, S. & Benoit, P. H. in *Chondrules and the Protoplanetary Disk* (eds Hewins, R. H., Jones, R. H. & Scott, E. R. D.) 221–231 (Cambridge Univ. Press, Cambridge, 1996).
- Jones, R. H. FeO-rich, porphyritic pyroxene chondrules in unequilibrated ordinary chondrites. *Geochim. Cosmochim. Acta* **60**, 3115–3138 (1996).
- Hewins, R. H. Chondrules. *Annu. Rev. Earth Planet. Sci.* **25**, 61–83 (1997).

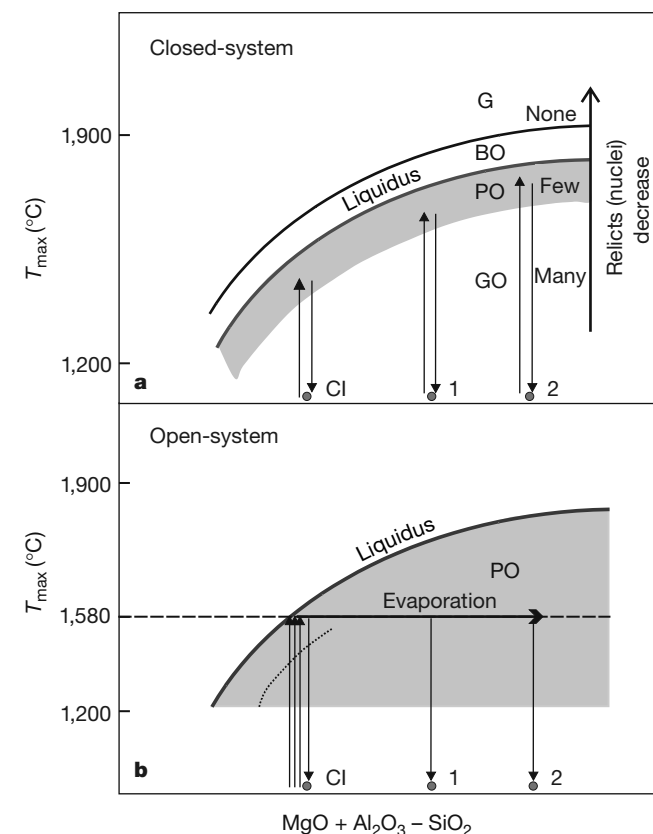


Figure 3 Effect of evaporation on the correlation between maximum temperature and chondrule bulk composition (MgO + Al₂O₃ – SiO₂ in wt%). This correlation, required for closed-system melting, is not necessary under evaporation. Field boundaries are shown for chondrule textures (G, glass; BO, barred olivine; PO, porphyritic olivine; GO, granular olivine). **a**, In closed-system melting, porphyritic texture is confined to maximum temperatures just below the liquidus, where the number of nuclei is sufficiently low to form large crystals upon cooling. Above the liquidus nuclei are very few (BO) or absent (G). At temperatures of 100 °C or more below the liquidus, many nuclei survive and grow into small grains (GO). In such a system producing porphyritic objects with compositions CI, 1 and 2 is possible only by using three different starting materials (CI, 1 and 2) and heating them to just the right temperature (0–50 °C below the liquidus). **b**, In experiments performed by open-system melting (at 1,580 °C) porphyritic textures were produced for an extremely wide range of (final) liquidus temperatures (~200 °C), as numbers of nuclei established close to the initial liquidus change little during evaporation. In such a system porphyritic objects with compositions CI, 1 and 2 are produced by the evaporation of a single starting composition (in this case CI) at a constant temperature for different lengths of time. Evaporation temperature is 0–50 °C below the liquidus of starting material. Arrows indicate heating and cooling paths. Compositions 1 and 2 correspond to type IIA and type IA, respectively.

6. Cuzzi, J. N., Dobrovolskis, A. R. & Hogan, R. C. in *Chondrules and the Protoplanetary Disk* (eds Hewins, R. H., Jones, R. H. & Scott, E. R. D.) 35–43 (Cambridge Univ. Press, Cambridge, 1996).
7. Boss, A. P. in *Chondrules and the Protoplanetary Disk* (eds Hewins, R. H., Jones, R. H. & Scott, E. R. D.) 257–263 (Cambridge Univ. Press, Cambridge, 1996).
8. Scott, E. R. D., Love, S. G. & Krot, A. N. in *Chondrules and the Protoplanetary Disk* (eds Hewins, R. H., Jones, R. H. & Scott, E. R. D.) 87–96 (Cambridge Univ. Press, Cambridge, 1996).
9. Hewins, R. H. & Radomsky, P. M. Temperature conditions for chondrule formation. *Meteoritics* **25**, 309–318 (1990).
10. Lofgren, G. E. in *Chondrules and the Protoplanetary Disk* (eds Hewins, R. H., Jones, R. H. & Scott, E. R. D.) 187–196 (Cambridge Univ. Press, Cambridge, 1996).
11. Grossman, J. N., Rubin, A. E., Nagahara, H. & King, E. A. in *Meteorites and the Early Solar System* (eds Kerridge, J. F. & Matthews, M. S.) 619–659 (Univ. Arizona Press, Tucson, 1988).
12. Hewins, R. H., Yu, Y., Zanda, B. & Bourot-Denise, M. Do nebular fractionations, evaporative losses, or both, influence chondrule compositions? *Antarct. Met. Res.* **10**, 294–317 (1997).
13. Nagahara, H., Kushiro, I. & Mysen, B. O. Evaporation of olivine: low pressure phase relations of the olivine system and its implication for the origin of chondritic components in the solar system. *Geochim. Cosmochim. Acta* **58**, 1951–1963 (1994).
14. Yu, Y. & Hewins, R. H. Transient heating and chondrule formation—evidence from Na loss in flash heating simulation experiments. *Geochim. Cosmochim. Acta* **62**, 159–172 (1998).
15. Anders, E. & Grevesse, N. Abundances of the elements: meteoritic and solar. *Geochim. Cosmochim. Acta* **53**, 197–214 (1989).
16. Wood, J. A. & Hashimoto, A. Mineral equilibrium in fractionated nebular systems. *Geochim. Cosmochim. Acta* **57**, 2377–2388 (1993).
17. Yu, Y. et al. Mass fractionation of K isotopes in chondrule evaporation experiments. *Lunar Planet. Sci. XXIX*, (1998).
18. Weinbruch, S., Palme, H. & Spettel, B. Refractory forsterite in primitive meteorites: Condensates from the solar nebula? *Meteor. Planet. Sci.* **35**, 161–171 (2000).
19. Alexander, C. M. O'D. in *Chondrules and the Protoplanetary Disk* (eds Hewins, R. H., Jones, R. H. & Scott, E. R. D.) 233–241 (Cambridge Univ. Press, Cambridge, 1996).
20. Jones, R. H. & Lofgren, G. E. A comparison of FeO-rich, porphyritic olivine chondrules in unequilibrated chondrites and experimental analogues. *Meteoritics* **28**, 213–221 (1993).
21. Weinbruch, S. & Müller, W. F. Constraints on the cooling rates of chondrules from the microstructure of clinopyroxene and plagioclase. *Geochim. Cosmochim. Acta* **59**, 3221–3231 (1995).
22. Weinbruch, S., Buttner, H., Rosenhauer, M. & Hewins, R. H. On the lower limit of chondrule cooling rates. *Meteor. Planet. Sci.* **33**, 65–74 (1998).
23. Walker, D., Jurewicz, S. & Watson, E. B. Adcumulus dunite growth in a laboratory thermal gradient. *Contrib. Mineral. Petrol.* **99**, 306–319 (1988).
24. Connolly, H. C. Jr & Hewins, R. H. Chondrules as products of dust collisions with totally molten droplets within a dust-rich nebular environment: an experimental investigation. *Geochim. Cosmochim. Acta* **59**, 3231–3246 (1995).

Acknowledgements

We thank S. Morse for starting materials and the Muséum National d'Histoire Naturelle in Paris for their support. This work was funded by NASA.

Correspondence and requests for materials should be addressed to B.C. (e-mail: bosmat@eden.rutgers.edu).

Quasicrystalline valence bands in decagonal AlNiCo

Eli Rotenberg*, W. Theis†, K. Horn‡ & P. Gille§

*MS2-400, Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

†Institut für Experimentalphysik der Freien Universität Berlin, Arnimallee 14, D-14195 Berlin, Germany

‡Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14195, Berlin, Germany

§Institut für Kristallographie und Angewandte Mineralogie der Ludwig-Maximilians-Universität, Theresienstrasse 41, D-80333, München, Germany

Quasicrystals are metallic alloys that possess perfect long-range structural order, in spite of the fact that their rotational symmetries are incompatible with long-range periodicity. The exotic structural properties of this class of materials¹ are accompanied by physical properties that are unexpected for metallic alloys. Considerable progress in resolving the geometric structures of quasicrystals has been made using X-ray and neutron diffraction, and concepts such as the quasi-unit-cell model² have provided

theoretical insights. But the basic properties of the valence electronic states—whether they are extended as in periodic crystals or localized as in amorphous materials—are still largely unresolved³. Here we investigate the electronic bandstructure of quasicrystals through angle-resolved photoemission experiments on decagonal Al_{71.8}Ni_{14.8}Co_{13.4}. We find that the *s-p* and *d* states exhibit band-like behaviour with the symmetry of the quasiperiodic lattice, and that the Fermi level is crossed by dispersing *d*-bands. The observation of free-electron-like bands, distributed in momentum space according to the surface diffraction pattern, suggests that the electronic states are not dominated by localization.

As a consequence of their translational symmetry, ordinary periodic crystals have electronic bands consisting of wavefunctions in the Bloch form, $\Psi_{\mathbf{k}} = u_{\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}$, where the plane wave with wavevector $\mathbf{k}$ is modulated by the lattice-periodic function $u_{\mathbf{k}}(\mathbf{r})$. (With the convention that $\hbar = 1$, electron wavevector $\mathbf{k}$ and crystal momentum $\mathbf{p} = \hbar\mathbf{k}$ are equivalent). The reciprocal $\mathbf{k}$ -space ordering is also periodic, with the bandstructure from the central Brillouin zone repeated in each primitive unit cell in $\mathbf{k}$ -space. In random systems, on the other hand, the translational symmetry is lost, $\mathbf{k}$ is no longer a good quantum number and disorder leads to the formation of localized eigenstates with an exponential decay of the amplitudes⁴. A fundamental, open question is whether valence states in quasicrystals exhibit band-like behaviour, with the concepts of crystal momentum and Brillouin zones adapted to aperiodic order^{5,6}, or whether they exhibit localized or “confined” states⁷ which one might expect in view of the cluster-like structure of quasicrystals.

The electronic structure of quasicrystals has been studied by means of electrical resistivity, magnetoresistivity, Hall-effect and thermopower experiments⁸, among others, and important information is gained from these measurements. The interpretation of such experiments in terms of the character of the electronic states is not straightforward, however. Angle-resolved photoemission, on the other hand, yields direct information on the nature and $\mathbf{k}$ -dependent dispersion of electronic states⁹; hence, this method is ideal for our search for the band-like character of quasicrystalline electronic states. An earlier report in fact found some evidence for such states in icosahedral AlPdMn (ref. 10). Instead, we chose Al_{71.8}Ni_{14.8}Co_{13.4} decagonal quasicrystals (henceforth AlNiCo) for our study because they possess both regular crystalline order (along the ten-fold axis)

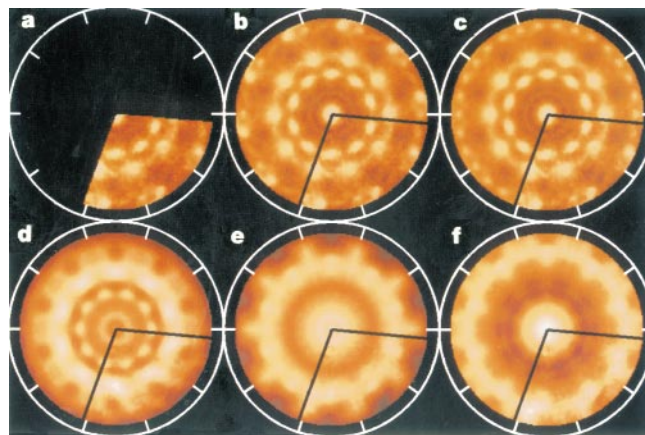


Figure 1 Angle-resolved photoemission intensity maps for AlNiCo valence-band electrons for various binding energies relative to the Fermi level E_F . Data were acquired for a photon energy of 130 eV. **a**, Raw data for electrons at E_F emitted into a 105° azimuth range. The radial coordinate is proportional to the polar angle of emission θ from 0° (centre) to 42° (outer rim). **b**, **c**, The same data as in **a** are compared to ten-fold rotational and subsequent mirror-plane averaging, respectively. **d–f**, Similarly processed data for electron energies 1.03, 1.84 and 2.77 eV below E_F , respectively.

and quasiperiodic order (normal to the ten-fold axis). This allows us to observe quasicrystalline and crystalline ordering in the same bands, by examining the k -space distribution of electrons moving within and out of the surface plane. AlNiCo crystals were grown by the Czochralski method¹¹ and samples with ten-fold surfaces were cut and polished after orientation by Laue X-ray diffraction. The surface was cleaned by ion bombardment followed by annealing in the ultra-high-vacuum chamber and was characterized by low energy electron diffraction (LEED), secondary electron imaging¹², and core-level photoelectron diffraction, all of which showed a ten-fold symmetric pattern characteristic of quasicrystalline order. Surface preparation and experiments were performed at beamline 7.0 at the Advanced Light Source, with a combined instrumental energy resolution of about 55 meV and an angular acceptance better than 1.4 degrees. Angle-resolved valence bands were recorded by rotating the sample polar (θ) and azimuthal (ϕ) angles with the analyser and light source orientations fixed¹³. By measuring the kinetic energy and angles of emission relative to the sample surface, we can, with simple assumptions, determine the three-dimensional momenta $\mathbf{k}$ of the electrons⁹.

That the valence electrons 'feel' the symmetry of the quasicrystalline structure can readily be inferred from a set of angular intensity maps of the valence electrons, representing the photoemission intensity emitted into a partial angular sector above the sample (Fig. 1a). The repeating elements in this image are separated by an azimuth angle of 36° , providing clear evidence of a ten-fold rotationally symmetric emission pattern. Experimental noise can therefore be reduced by a ten-fold rotational averaging of the data (Fig. 1b). Because of the arbitrary orientation of the light's polarization vector to our sample surface, the photoemission data show a slight breaking of mirror-plane symmetry that changes the intensities (but not the positions) of features across mirror planes. We can restore the intrinsic mirror-plane symmetry, as shown by

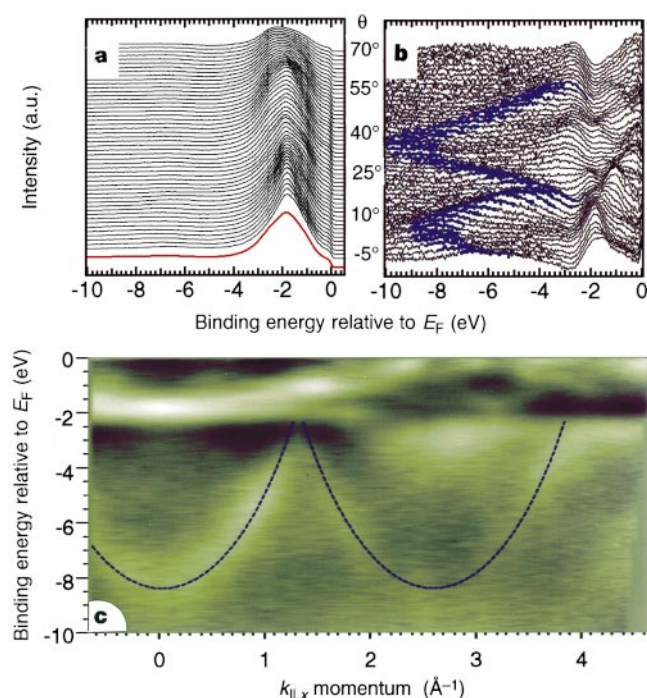


Figure 2 Angle-resolved photoemission spectra for AlNiCo valence electrons for photon energy 95 eV. **a**, A series of photoemission spectra for various polar angles θ . The red spectrum at the bottom is the angle-integrated spectrum. **b**, The same data as in **a**, normalized to the angle-integrated spectrum. The states indicated in blue, weak in **a**, are now apparent. **c**, An intensity plot of the same data, now in the form of a conventional band map, obtained after mapping the angle θ into momentum $k_{\parallel,x}$. The dashed line indicates the parabolic band-like character of the s - p -derived states.

electron diffraction data (Fig. 3a), by a further mirror averaging (Fig. 1c).

The data show a complex pattern. The least-bound electrons at the Fermi energy (E_F) (Fig. 1a–c) display various sharp maxima at specific emission angles which appear to evolve continuously as deeper energies are sampled (Fig. 1d–f). The evolution of these patterns, occurring rapidly over just a few electron volts, suggests that the electronic states disperse continuously with momentum $\mathbf{k}$. Such rapid changes over a narrow energy range rule out an explanation of these patterns in terms of elastic scattering of the electrons after the photoemission event¹⁴.

To prove that the electronic energies disperse continuously with $\mathbf{k}$, we acquired valence band spectra versus momentum parallel ($k_{\parallel}$) to the ten-fold symmetric sample surface. Figure 2a shows a set of spectra recorded at different polar emission angles θ along $k_{\parallel,x}$, one of the two-fold axes in the $k_{\parallel}$ -plane. The spectra are dominated by a broad peak at a binding energy of around 2 eV, due mostly to the Co and Ni d -electrons, but additional, weaker contributions from Al, Ni and Co s - p electrons are also present down to about 8.5 eV below E_F . (The s - p states appear weak merely as a consequence of their low relative photoemission cross-sections.) Deviations from the angle-averaged spectrum (red line) clearly demonstrate a $\mathbf{k}$ -dependence of the d -derived spectral features.

In order to raise the s - p states to the same scale as the d -states, we have divided the angle-resolved spectra by the angle-averaged spectrum and plotted the resulting normalized data as individual spectra in Fig. 2b and as an intensity map in Fig. 2c. The normalized data in Fig. 2b–c reveal two new bands, assigned to s - p -derived states, which strongly disperse at a binding energy of 3–8.5 eV; they have an almost free electron-like dispersion as their energy follows the simple parabolic form $E = \hbar^2 k^2 / 2m^*$, with an effective mass m^* of about 0.9 electron masses. The changes in the d -level region are difficult to assign, but we can identify at least one band-like feature,

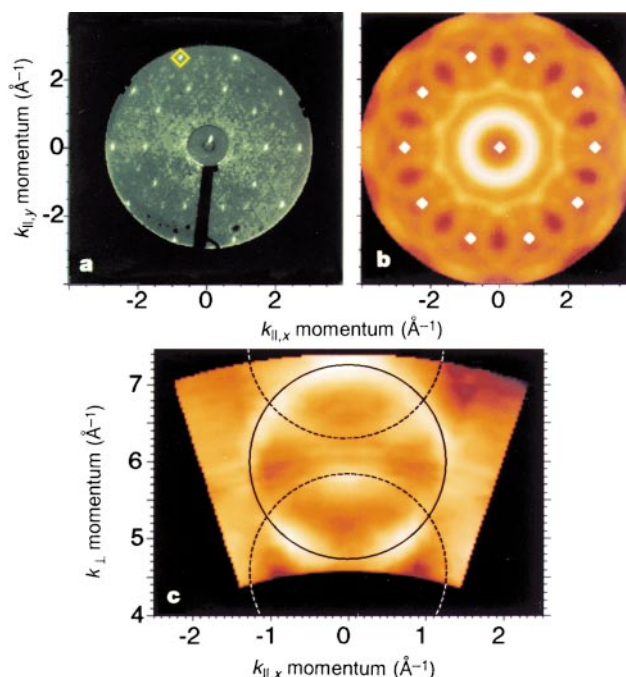


Figure 3 In-plane and out-of-plane order in $\mathbf{k}$ -space. **a**, Low-energy electron diffraction pattern for 71.5 eV electrons, showing the ten-fold, in-plane aperiodic order. **b**, Emission profile of 6 eV binding energy electrons, for a wide range of in-plane electron momenta $k_{\parallel}$. A central ring and ten outer rings are observed, establishing that the s - p electrons sample the aperiodic order. The brightest, outermost ring of spots in **a** are reproduced in **b** by the diamonds. **c**, A vertical cut through the central spherical constant-energy surface shows a strong central ring (solid) with weaker circles spaced equally above and below it (dashed).

with its highest binding energy at normal emission ($\theta = 0^\circ$, $\mathbf{k}_\parallel = 0$), that turns upwards and appears to cross E_F at about $k_\parallel = 2.5 \text{ \AA}^{-1}$. The complete map of such crossings, of which Fig. 2c is just a partial sampling, forms the Fermi surface, whose topology is important for predictions of anomalous transport properties¹⁵ and may also contribute to the stability of quasicrystals.

We can use the simple parabolic dispersion of the s - p states to observe the quasiperiodic ordering of the bands in $\mathbf{k}$ -space. In Fig. 3a, we show a ten-fold symmetric LEED diffraction pattern from our sample for an electron kinetic energy of 71.5 eV. The LEED pattern emerges from scattering at the atomic cores and therefore derives from the aperiodic geometry. The diffraction pattern from a quasicrystalline surface in principle forms a set that is dense in $\mathbf{k}$ -space, that is, there is an unlimited number of "reciprocal lattice vectors" and thus an unlimited number of diffraction spots¹⁶. However, as is obvious from the pattern in Fig. 3, there is a set of spots with dominant intensities that may be taken to construct a "quasi-Brillouin zone" if we consider only the dominant Fourier components of the lattice potential^{5,6}. Similar to the case of periodic crystals, we may expect to observe bands from the centre of the quasi-Brillouin zone to be replicated to the positions of the principal (and possibly weaker) diffraction spots.

To test this idea, we have sampled the four experimental degrees of freedom (binding energy E , in-plane momentum $\mathbf{k}_\parallel = (k_{\parallel,x}, k_{\parallel,y})$ and out-of-plane momentum, $k_\perp$) in another way. Since the free-electron energy is $E = \hbar^2 k^2 / 2m^*$, then for constant energy the electron momenta are restricted to spherical surfaces in three-dimensional $\mathbf{k}$ -space. Circular, cross-sectional cuts through these spheres can be visualized by sampling the $\mathbf{k}$ -dependent intensity distribution of electrons (with energy 6 eV below E_F) in two complementary ways. In Fig. 3b, we show data for electrons at constant $|\mathbf{k}|$, sampling states over a wide range of $\mathbf{k}_\parallel$. In Fig. 3c we show states in the $k_{\parallel,x}$ - $k_\perp$ plane, thus sampling over a wide range of $k_\perp$ near $\mathbf{k}_\parallel = 0$.

In Fig. 3b, we observe a circular ring around the centre ($\mathbf{k}_\parallel = 0$) and ten outlying rings of equal diameter. The central and the ten outer rings correspond to 6 eV level crossings of the bands from Fig. 2c centred at $k_{\parallel,x} = 0$ and about $2.5 \text{ \AA}^{-1}$, respectively. Furthermore, we find that the ten outer rings are centred on each of the dominant diffraction spots in Fig. 3a. Therefore, the s - p band emerging from the central quasi-Brillouin zone has its counterparts in identical bands distributed aperiodically in $\mathbf{k}$ -space. Their arrangement is easily understood in terms of a repeated zone scheme owing to the influence of the in-plane aperiodic potential.

As there is a periodic arrangement of atomic planes normal to the sample surface, we expect a periodic repetition of bands in this direction. This is readily observed in the data in the $k_{\parallel,x}$ - $k_\perp$ plane (Fig. 3c). Ring-like contours are again observed, corresponding to the $\mathbf{k}$ -space unit cells along the ten-fold axis; however, in contrast to the quasiperiodic ordering of the bands with its ten-fold symmetry forbidden in periodic crystals, the ring-like features here are arranged equidistantly. This arrangement continues for at least four complete periods along $k_\perp$ for both the s - p and the d -bands. The distance between rings, $1.55 \text{ \AA}^{-1}$, corresponds to an interlayer distance of $d = 2\pi / 1.55 \text{ \AA}^{-1} = 4.05 \text{ \AA}$, in good agreement with X-ray diffraction measurements ($4.08 \text{ \AA}$, ref. 17). This demonstrates that periodic ordering prevails in the direction parallel to the ten-fold axis, as expected from the structure of AlNiCo.

How do the electronic wavefunctions in a quasicrystal compare to those in a periodic metal? The diffraction pattern consists of a few dominant spots, suggesting that the atomic distribution (and hence the electrostatic potential entering the Schrödinger equation) can be expanded in terms of an aperiodic lattice in $\mathbf{k}$ -space. The solutions to the Schrödinger equation will therefore be of the form $\Psi_{\mathbf{k}} = u_{\mathbf{k}}(r)e^{i\mathbf{k}\cdot\mathbf{r}}$, which is similar to the Bloch form for conventional crystals. The main difference between quasicrystals and periodic crystals is that in the latter $u_{\mathbf{k}}(r)$ is a Fourier series over

the periodic reciprocal lattice and therefore has the full real-space periodicity of the solid, while in the present case $u_{\mathbf{k}}(r)$ is a summation over the countably dense, aperiodic reciprocal lattice. This appears to be a useful expansion³ since only a few special reciprocal lattice points dominate the spectra, consistent with our observation of parabolic bands centred on the same reciprocal lattice points as the principal electron diffraction spots^{5,6}.

The photoemission data are important for characterizing the electronic states of quasicrystals. The eigenvalue spectrum of a system can be classified as continuous or discrete, corresponding to extended or localized eigenstates. Continuous spectra are characteristic for periodic systems, where Bloch's theorem holds, but in random systems such as amorphous metals, a discrete spectrum occurs, and the states are localized with an exponential decay of the amplitudes⁴. We have shown that the localization of states in $\mathbf{k}$ -space appears sufficient to ensure that at least some states are extended in real space (although the possibility remains that there are additional localized states present). Furthermore, the wide bandwidths (2–8 eV) and small effective masses (comparable to free electrons) indicate a high degree of extended character both within and out of the quasicrystalline planes.

In fact, quasicrystals may be an intermediate case, as shown by renormalization group studies of one-dimensional quasiperiodic chains, in that they have critical states characterized by a power-law decay of their amplitudes¹⁸. These states are derived from states localized to a cluster of size d which can resonantly tunnel to identical nearby clusters. These neighbouring clusters are separated by a distance $2d$ or less according to Conway's theorem¹⁹, which is apparently satisfied for AlNiCo (ref. 2). Critical eigenstates in quasicrystals thus derive their properties from both limiting cases, that is, a dispersion relation according to the potential distribution of the local environment, but with amplitudes that are strongly dampened. Our data cannot rule out the possibility that the eigenstates are critical; in fact the dampening of such states may explain the weakness of features in our spectra—both s - p - and d -derived—when compared to band-like emission from periodic crystals. □

Received 15 March; accepted 2 June 2000.

- Schechtman, D., Blech, I., Gratias, D. & Cahn, J. W. Metallic phase with long-range orientational order and no translational symmetry. *Phys. Rev. Lett.* **53**, 1951–1953 (1984).
- Steinhardt, P. J. *et al.* Experimental verification of the quasi-unit-cell model of quasicrystal structure. *Nature* **396**, 55–57 (1998); correction *Nature* **399**, 84 (1999).
- Jano, C. *Quasicrystals: A Primer* 359–399 (Oxford Univ. Press, New York, 1997).
- Hafner, J. & Krajci, M. in *Physical Properties of Quasicrystals* (ed. Stadnik, Z. M.) (Springer, Berlin, 1999).
- Smith, A. P. & Ashcroft, N. W. Pseudopotentials and quasicrystals. *Phys. Rev. Lett.* **59**, 1365–1368 (1987).
- Niizeki, K. & Akamatsu, T. Special points in the reciprocal space of an icosahedral quasi-lattice and the quasi-dispersion relation of electrons. *J. Phys. Cond. Mat.* **2**, 2759–2771 (1990).
- Trambly de Laissardiere, G. & Mayou, D. Clusters and localization of electrons in quasicrystals. *Phys. Rev. B* **55**, 2890–2893 (1997).
- Rapp, Ö. in *Physical Properties of Quasicrystals* (ed. Stadnik, Z. M.) 127–167 (Springer, Berlin, 1999).
- Kevan, S. D. (ed.) *Angle Resolved Photoemission* (Elsevier, Amsterdam, 1992).
- Wu, X. *et al.* Electronic band dispersion and pseudogap in quasicrystals—angular-resolved photoemission studies on icosahedral $\text{Al}_{70}\text{Pd}_{13}\text{Mn}_{15}$. *Phys. Rev. Lett.* **75**, 4540–4543 (1995).
- Gille, P., Dreier, P., Gräber, M. & Scholpp, T. Large single-grain AlCoNi quasicrystals grown by the Czochralski method. *J. Cryst. Growth* **207**, 95–101 (1999).
- Zurkirch, M., Bolliger, B., Erbuda, M. & Kortan, A. R. Structural transformations at the surface of the decagonal quasicrystal $\text{Al}_{70}\text{Co}_{15}\text{Ni}_{15}$. *Phys. Rev. B* **58**, 14113–14116 (1998).
- Rotenberg, E. & Kevan, S. D. Evolution of Fermi level crossings versus H coverage on W(110). *Phys. Rev. Lett.* **80**, 2905–2908 (1998).
- Osterwalder, J., Greber, T., Hüfner, S. & Schlappbach, L. X-ray-photoelectron diffraction from a free-electron-metal valence band: evidence for hole-state localization. *Phys. Rev. Lett.* **64**, 2683–2686 (1990).
- Roche, S. & Fujiwara, T. Fermi surfaces and anomalous transport in quasicrystals. *Phys. Rev. B* **58**, 11338–11343 (1998).
- Senchal, M. *Quasicrystals and Geometry* (Cambridge Univ. Press, Cambridge, 1995).
- Steurer, W., Haibach, T., Zhang, B., Kek, S. & Lück, R. The structure of decagonal $\text{Al}_{70}\text{Ni}_{15}\text{Co}_{15}$. *Acta Crystallogr. B* **49**, 661–675 (1993).
- Fujiwara, T., Kohmoto, M. & Tokihiro, T. Multifractal wave functions on a Fibonacci lattice. *Phys. Rev. B* **40**, 7413–7416 (1989).
- Gardner, M. Extraordinary nonperiodic tiling that enriches the theory of tiles. *Sci. Am.* **236**, 110–117 (1977).

Acknowledgements

This work is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, of the US Department of Energy at Lawrence Berkeley National Laboratory. P.G. and W.T. acknowledge support from the Deutsche Forschungsgemeinschaft.

Correspondence and requests for materials should be addressed to E.R. (e-mail: erotenberg@lbl.gov).

A DNA-fuelled molecular machine made of DNA

Bernard Yurke*, Andrew J. Turberfield*†, Allen P. Mills Jr*, Friedrich C. Simmel* & Jennifer L. Neumann*

* Bell Laboratories, Lucent Technologies, 600 Mountain Avenue, Murray Hill, New Jersey 07974, USA

† Department of Physics, University of Oxford, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK

Molecular recognition between complementary strands of DNA allows construction on a nanometre length scale. For example, DNA tags may be used to organize the assembly of colloidal particles^{1,2}, and DNA templates can direct the growth of semiconductor nanocrystals³ and metal wires⁴. As a structural material in its own right, DNA can be used to make ordered static arrays of tiles⁵, linked rings⁶ and polyhedra⁷. The construction of active devices is also possible—for example, a nanomechanical switch⁸, whose conformation is changed by inducing a transition in the chirality of the DNA double helix. Melting of chemically modified DNA has been induced by optical absorption⁹, and conformational changes caused by the binding of oligonucleotides or other small groups have been shown to change the enzymatic activity of ribozymes^{10–13}. Here we report the construction of a DNA machine in which the DNA is used not only as a structural material, but also as ‘fuel’. The machine, made from three strands of DNA, has the form of a pair of tweezers. It may be closed and opened by addition of auxiliary strands of ‘fuel’ DNA; each cycle produces a duplex DNA waste product.

Single strands of DNA composed of complementary sequences of the bases adenine, cytosine, guanine and thymine (A, C, G and T) hybridize to form a stable duplex (double helix) bound together by hydrogen bonds between complementary base pairs (A–T and C–G). Our machine is prepared by mixing stoichiometric quantities of three strands, A, B and C, in SPSC buffer (see Methods) at 20 °C to give a final concentration of 1 μ M; the base sequences chosen for the three strands are given in Fig. 1. The structure and operation of the machine are shown in Fig. 2. Strand A consists of two 18-base sequences which hybridize with complementary

sequences at the ends of strands B and C to form two stiff¹⁴ arms; the hinge is formed from a four-base single-stranded region of A between the regions hybridized to strands B and C. In the machine’s rest state, the remaining unhybridized 24-base portions of the 42-base strands B and C dangle floppily from the ends of the tweezers: double-stranded DNA has a persistence length of the order of 100 base pairs^{14,15}, whereas at 1 M salt concentration single-stranded DNA has a persistence length of about 1 nm (ref. 16)—or approximately three bases.

Strand A is labelled at the 5’ and 3’ ends with dyes TET (5’ tetrachloro-fluorescein phosphoramidite) and TAMRA (carboxy-tetramethylrhodamine), respectively. When TET is excited by the 514.5-nm emission of an argon ion laser, it fluoresces with a peak emission wavelength of 536 nm; this emission is quenched by resonant intramolecular energy transfer from TET to TAMRA (a longer-wavelength dye whose absorption band overlaps the emission band of TET) with an efficiency that decreases rapidly as the distance between the dyes increases¹⁷. Fluorescence quenching is used as an indicator¹⁸ to titrate strands B and C against A; as half of A is straightened from a random coil by hybridization with B (or C), the mean separation between the dye groups on A increases leading to a fivefold increase in the fluorescence intensity. The cumulative effect of hybridization with both B and C is a sevenfold increase in fluorescence intensity in the rest state.

The assembled tweezers are opened and closed with fuel strands F and $\bar{F}$. The 56-base closing strand F consists of two consecutive 24-base sections, which are complementary to the dangling ends of B and C, with an additional 8-base overhang section. Figure 2b shows how closing strand F hybridizes with the free ends of strands B and C, pulling the ends of the tweezers together. The average free-energy change associated with the hybridization of a complementary base pair is -78 meV (-1.8 kcal mol⁻¹) at 20 °C (ref. 19) and the separation between base pairs in single-stranded DNA is 0.43 nm (ref. 20), giving an average closing force of about 15 pN which is consistent with that required to pull apart double-stranded DNA²¹. This is at the upper end of the range of measured forces exerted by single-group kinesin^{22,23} and myosin^{24,25} motors. The second component of the fuel is removal strand $\bar{F}$, the complement of F: the additional free energy gained when the overhang (single-stranded in the initial state) hybridizes with $\bar{F}$ ensures that when a stoichiometric quantity of $\bar{F}$ is added it removes F from the machine to form a double-stranded waste product $F\bar{F}$ and returns the tweezers to the open state. Hybridization between the closing and removal fuel strands is expected to occur first at the exposed overhang²⁶ and to proceed by branch migration²⁷, a random walk of the junction between the region of F newly hybridized to the removal strand $\bar{F}$ and the region still hybridized to the tweezer components B and C; this branch migration continues until both B and C have been completely displaced and the $F\bar{F}$ duplex diffuses away. The random walk occurs sufficiently quickly that the rate-limiting step in such strand-displacement reactions is the endothermic nucleation of a region in which complementary bases are joined^{26,28,29} (in this case, the nucleation of hybridization between the F and $\bar{F}$ strands): this

Name	Domain 1	Sequence	Domain 2
A	5' TGCCTTGAAGAGCGACCAT	CAACCTGGAATGCTTCGGAT 3'	
B	5' GGTGCTCTTACAAGGCA	CTGGTAACAATCACGGTCTATGCG 3'	
C	5' GGAGTCTCTACTGTCTGAAGTAAAG	ATCCGAAGCATTCCAGGT 3'	
F	5' CGCATAGACCGTGATTGTTACCAAG	CGTTAGTTCAGACAGTAGGACTCC TGCTACGA 3'	
β	5' GGTGCTCTTACAAGGCA	CAGCTAGTTTCAAGTGGCAAGTC 3'	
γ	5' GCAGGCTTCTACATATCTGACGAG	ATCCGAAGCATTCCAGGT 3'	
$F_{B\gamma}$	5' CGCATAGACCGTGATTGTTACCAAG	CTCGTCAGATATGTAGAAGCCTGC ACGTCGAT 3'	
$F_{C\beta}$	5' GACTTGCCACTGTGAACTAGCTG	CGTTAGTTCAGACAGTAGGACTCC TGTCAGA 3'	

Figure 1 Oligonucleotide sequences. Blue and green colouring indicates sections of B and C that hybridize to complementary sections of closing strand F; the red section of F is the overhang at which interaction with opening strand $\bar{F}$ begins. Shorter oligonucleotides corresponding to the domains indicated are used in a control experiment

described in the text. Oligonucleotides B1 and C2 used in the control experiment are complementary to domains A1 and A2 respectively of A, including the central four bases that form the hinge, and are therefore two bases longer than the sequences shown above.

is consistent with our observation that the measured second-order rate constants for opening and closing the tweezers are approximately equal.

Dye quenching is used to titrate the closing and removal strands and to determine the state of the machine. The fluorescence intensity drops by a factor of six when the tweezers are closed, back to approximately the same level as from unhybridized, randomly coiled A strands. Figure 3 shows the fluorescence intensity variation as the tweezers are cycled seven times between the open and closed states by successive additions of strands F and $\bar{F}$. The switching time for the machine is approximately 13 s.

To calibrate the motion of the device, we measure the fluorescence intensity as a function of the separation between TET and TAMRA groups. A 24-base oligonucleotide labelled at the 5' end with TET and a separate 24-mer labelled at the 3' end with TAMRA are hybridized to complementary sequences at either end of a $(48+n)$ -mer ($n = 10, 15, 20, 40$), whose central n -base spacer section is straightened by hybridization to a complementary n -mer. The dye-labelled ends of the two 24-mers are separated only by the spacer section; its length is much less than the persistence length of double-stranded DNA, so it acts as a rigid rod holding the dye groups at a fixed separation. The fluorescence intensity from the open tweezers (an average over many possible open configurations) is equal to the interpolated value for a 17-base spacer: we deduce an approximate average value for the difference in separation between the ends of the tweezers in the open and closed states of 6 nm, corresponding to an angle between the open tweezer arms of 50° .

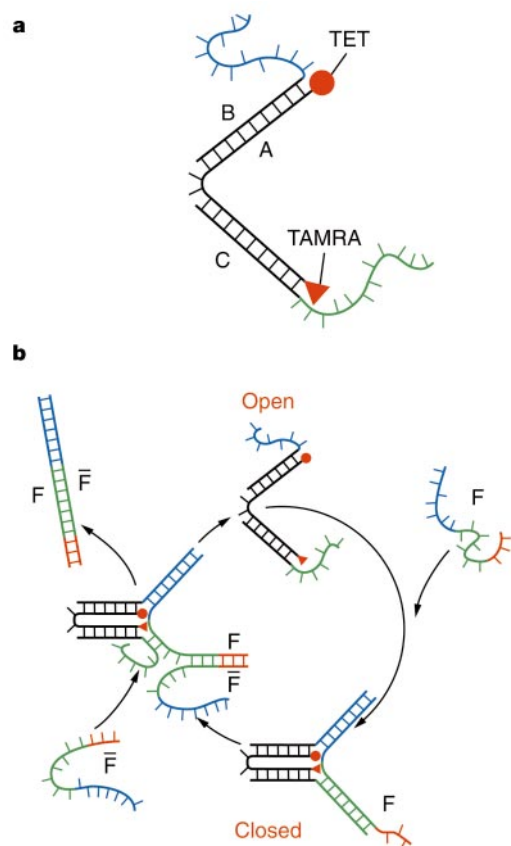


Figure 2 Construction and operation of the molecular tweezers. **a**, Molecular tweezer structure formed by hybridization of oligonucleotide strands A, B and C. **b**, Closing and opening the molecular tweezers. Closing strand F hybridizes with the dangling ends of strands B and C (shown in blue and green) to pull the tweezers closed. Hybridization with the overhang section of F (red) allows $\bar{F}$ strand to remove F from the tweezers, forming a double-stranded waste product FF and allowing the tweezers to open. Complementary sections of B, C, F and $\bar{F}$ that hybridize to close and open the tweezers are coloured as in Fig. 1.

As a control we add closing strand F to tweezers which incorporate a modified strand A labelled with TET only: we observe only a 17% drop in the TET fluorescence intensity, five times less than in the case when the other end of the tweezers is labelled with TAMRA. When a fully labelled strand A (incorporating both dye groups) is used but one of the tweezer components B, C is left out, so that the tweezers cannot close, then changes in fluorescence intensity due to successive additions of F and $\bar{F}$ are less than 8% of the change achieved with the complete structure. In contrast, when tweezers are prepared with a fully labelled strand A which is shortened by one or two bases at the 5' end, changing the position of the point of attachment of the TET group but allowing the tweezers to close normally, we observe quenching ratios of 7.2 and 7.8 respectively on adding F. These experiments indicate that the quenching of TET emission on addition of the closing fuel strand is largely due to the proximity of the TAMRA group in the closed state, and is relatively insensitive to other changes in its environment.

In our model of 'properly closed' tweezers, shown schematically in Fig. 2b, the stiff arms of the tweezers are held together by the closing strand and quenching is due to the proximity of the two dye groups attached to a single strand of DNA. An alternative closed configuration—in which TET emission is also quenched—is a dimer or larger complex (referred to collectively as dimers below) in which closing strands tie the TET group on one strand close to the TAMRA on another (see inset to Fig. 4). To investigate dimer formation, we have made use of modified tweezers made from strands A, β , γ (Fig. 1). Strands β , γ have the same 18-base sequences complementary to A as strands B, C but have different 24-base free ends. Strands $F_{B\gamma}$, $F_{C\beta}$ cannot close either tweezer structure, but can hybridize to link free ends of strands B to those of strand γ , and free ends of strands C to those of strands β , joining one of the original tweezers to one of the modified tweezers to form a dimer.

We have used polyacrylamide gel electrophoresis to compare 'closed tweezers' $(A,B,C) + F$ with 'dimers' $(A,B,C) + (A,\beta,\gamma) + F_{B\gamma} + F_{C\beta}$. Figure 4 shows an image of the gel recorded using laser-excited dye fluorescence. Lanes a–e are controls containing incomplete structures; lane l contains dimers. Lanes f–k contain open (f,h,j) and closed (g,i,k) tweezers corresponding to successive additions of closing and removal fuel strands F and $\bar{F}$. Saturated bands in lanes b–f,h,j correspond to structures in which quenching of TET fluorescence is minimum. The dominant 'properly closed' band in the 'closed tweezer' lanes is absent from the 'dimer' lane, consistent with our interpretation that the dominant component of 'closed tweezers' does indeed correspond to 'properly closed' tweezers as designed. Low-intensity bands in the 'closed tweezer' lanes match bands in the 'dimer' lane l, confirming that some dimer formation does occur. To estimate the yield of 'properly closed' tweezers, we measure fluorescence quenching in mixtures of dye-labelled and unlabelled tweezers in which the fraction of TET and TAMRA groups bound together on adding closing fuel strands, and thus the degree of quenching depends on the proportion of 'properly closed' tweezers formed. These controls are described in the Supplementary Information: we estimate that about 80% of tweezers in the experiment recorded in Fig. 3 are 'properly closed' by the closing fuel strand F. We note that dimer formation could in principle be avoided completely by tethering the tweezers to a solid substrate to prevent interaction.

By comparing 'open tweezer' lanes f,h,j in Fig. 4 we deduce that, after each cycle of closing and opening, strand A remains intact and the initial configuration of the open tweezers is recovered. To test whether the duplex structures formed by hydrogen bonding between A, B and A, C survive the forces exerted by hybridization with F in the closed state, we have used an additional strand $\bar{A}$, the complement of A, to test the integrity of the machine by displacing B and C from A. In the absence of other strands, $\bar{A}$ hybridizes with A (producing an eight-fold increase in fluorescence intensity) with a

time to half-completion of 17 s for $[A] = [\bar{A}] = 1 \mu\text{M}$. When the same concentration of $\bar{A}$ is introduced to ready-formed tweezers in the open state, the time to half-completion for interaction is increased to $4.8 \times 10^3 \text{ s}$ by the need for thermally activated displacement of parts of B and C from A before hybridization with $\bar{A}$ can begin. The interaction of $\bar{A}$ with tweezers held in the closed state is an order of magnitude slower than in the open state (time to half-completion, $5 \times 10^4 \text{ s}$); the closed structure inhibits the winding and unwinding of strands that accompanies strand exchange. The high stability of the tweezers, even in the closed state where the hinge region may be strained, indicates that its structural integrity is maintained.

As a further test of our design we checked that the components of the machine interact as intended. We divided each of the strands $\{A, B, C, F, \bar{F}\}$ into two functionally separate domains, as indicated in Fig. 1. This separation of function may be understood with reference to Fig. 2: domains B1 and B2 of strand B are designed to hybridize with domains A1 and F1, and so on. Polyacrylamide gel electrophoresis at 20°C (see Methods) of stoichiometric mixtures of all pairs of oligonucleotides from the set $\{A, B, C, F, \bar{F}, A1 \dots \bar{F}1, A2 \dots \bar{F}2\}$ shows that only those pairs that are designed to hybridize do so. This demonstrates that there are no undesired interactions between components of the machine.

A similar structure to that reported here could be used to investigate the interaction between chemically active components attached to the ends of the tweezers or, if combined with a DNA

cage⁵, to alternately hide and reveal a target group. It is possible to imagine free-running machines using metastable DNA loops as fuel³⁰. Because the binding between fuel and machine is sequence-specific, the DNA strands that act as fuel may also serve as information carriers to coordinate components of a complex machine or to carry signals between machines. The use of DNA fuel allows precise control of movements on a nanometre length scale without prodding by a scanning microscope tip; it makes possible coordinated motion of devices which may incorporate elements chosen for their biochemical, optical or electrical properties. Controlled motion provides new opportunities in the design of nanostructures. □

Methods

Oligonucleotides were supplied by Integrated DNA Technologies, Inc.; strand A was purified by high-performance liquid chromatography, other strands were purified by polyacrylamide gel electrophoresis. Stock solutions were prepared by resuspending the lyophilized oligonucleotides in TE buffer (10 mM Tris pH 8.0, 1 mM EDTA) at a nominal $25 \mu\text{M}$ concentration. Fluorescence quenching was used as an indicator to titrate strands B, C against A and strands F, $\bar{F}$ against open tweezers (formed of stoichiometric quantities of A, B and C) in SPSC buffer (50 mM Na_2HPO_4 pH 6.5, 1 M NaCl). TET fluorescence was excited with the 514.5-nm line of an argon ion laser (mechanically chopped at 130 Hz), selected by an interference filter with bandpass 10 nm centred at 540 nm and detected by a Si photodiode and phase-sensitive detector. We estimate an error of 10% in determining relative concentrations. Absolute concentrations were obtained from ultraviolet absorption measurements by the commercial supplier of the oligonucleotides; from our measurements of relative concentrations, we estimate that the error in absolute concentrations is of the order of 30%. To form open tweezers 2 μl of stock solution of strand A was diluted to $1 \mu\text{M}$ using 48 μl SPSC buffer, then stoichiometric quantities of the stock

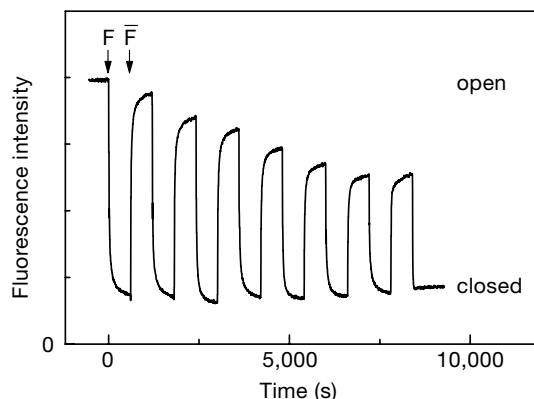


Figure 3 Cycling the molecular tweezers. By adding stoichiometric quantities of closing and removal strands F and $\bar{F}$ in sequence, the tweezers may be closed and opened

repeatedly. When the tweezers are closed, resonant energy transfer from the TET dye to the TAMRA quencher group reduces the fluorescence intensity.

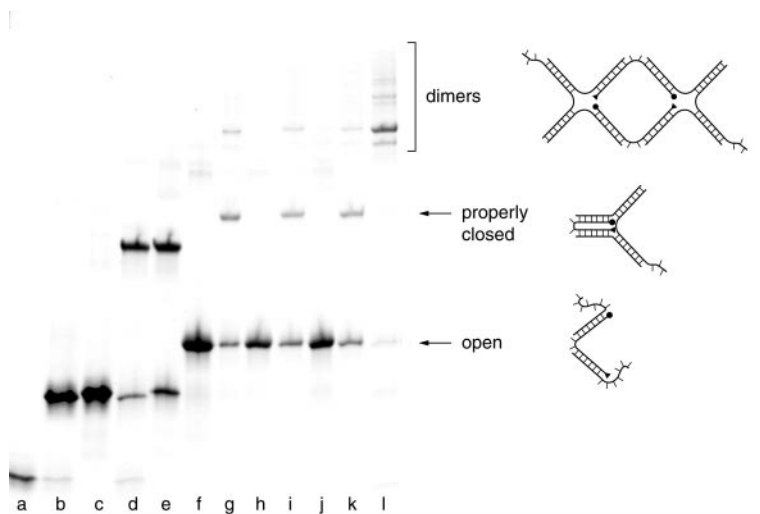


Figure 4 Analysis of tweezer formation by polyacrylamide gel electrophoresis. Lanes as follows: a, strand A; b, A+C; c, A+B; d, A+C+F; e, A+B+F; f, A+B+C (open tweezers); g, (A+B+C)+F (closed tweezers); h–k, two further cycles of opening and closing produced

by successive additions of $\bar{F}$ and F to closed tweezers; l, $(A,B,C)+(A,B,\gamma)+F_{B\gamma}+F_{C\beta}$ (only dimers can form). Schematic structures of open and properly closed tweezers and of a dimer complex are shown next to the appropriate bands.

solutions of strands B and C were added; the mixture was left for >20 min to allow the reaction to reach near-completion. Each addition of a stoichiometric quantity of a stock solution of strand F or F' resulted in ~3% further dilution of the reactants; after each addition (Fig. 3), reactants were mixed by rapid pipetting up and down lasting <8 s. All measurements were performed at 20 °C.

Polyacrylamide gel electrophoresis (Fig. 4) was performed using a Multiphor II flatbed electrophoresis system with an ExcelGel 48S precast gel (total acrylamide concentration $T = 12.5\%$, bisacrylamide concentration $C = 2\%$; Amersham Pharmacia Biotech). Reactions were performed and 5 μ l samples loaded in SPSC buffer with an approximate DNA concentration of 5 μ M for each species except $F_{B'}$, $F_{C'}$ (2.5 μ M). Each reaction was allowed a minimum of 20 min to proceed to near-completion. Fluorescence was excited by the 514.5-nm line of an argon ion laser, selected by an interference filter with bandpass 10 nm centred at 540 nm and recorded by a CCD camera; Fig. 4 is a composite of four images.

Received 12 May; accepted 7 June 2000.

1. Mirkin, C. A., Letsinger, R. L., Mucic, R. C. & Storhoff, J. J. A DNA-based method for rationally assembling nanoparticles into macroscopic materials. *Nature* **382**, 607–609 (1996).
2. Alivisatos, A. P. *et al.* Organization of 'nanocrystal groups' using DNA. *Nature* **382**, 609–611 (1996).
3. Coffey, J. L. *et al.* Dictation of the shape of mesoscale semiconductor nanoparticle assemblies by plasmid DNA. *Appl. Phys. Lett.* **69**, 3851–3853 (1996).
4. Braun, E., Eichen, Y., Sivan, U. & Ben-Yoseph, G. DNA-templated assembly and electrode attachment of a conducting silver wire. *Nature* **391**, 775–778 (1998).
5. Winfree, E., Liu, F., Wenzler, L. A. & Seeman, N. C. Design and self-assembly of two-dimensional DNA crystals. *Nature* **394**, 539–544 (1998).
6. Mao, C., Sun, W. & Seeman, N. C. Assembly of Borromean rings from DNA. *Nature* **386**, 137–138 (1997).
7. Chen, J. & Seeman, N. C. Synthesis from DNA of a group with the connectivity of a cube. *Nature* **350**, 631–633 (1991).
8. Mao, C., Sun, W., Shen, Z. & Seeman, N. C. A nanomechanical device based on the B-Z transition of DNA. *Nature* **397**, 144–146 (1999).
9. Asanuma, H., Ito, T., Yoshida, T., Liang, X. & Komiyama, M. Photoregulation of the formation and dissociation of a DNA duplex by using the *cis-trans* isomerization of azobenzene. *Angew. Chem. Int. Edn Engl.* **38**, 2393–2395 (1999).
10. Porta, H. & Lizardi, P. M. An allosteric hammerhead ribozyme. *Biotechnology* **13**, 161–164 (1995).
11. Tang, J. & Breaker, R. R. Rational design of allosteric ribozymes. *Chem. Biol.* **4**, 453–459 (1997).
12. Araki, M., Okuno, O., Hara, Y. & Sugiura, Y. Allosteric regulation of a ribozyme activity through ligand-induced conformational change. *Nucleic Acids Res.* **26**, 3379–3384 (1998).
13. Robertson, M. P. & Ellington, A. D. *In vitro* selection of an allosteric ribozyme that transduces analytes to amplicons. *Nature Biotechnol.* **17**, 62–66 (1999).
14. Smith, S. B., Finzi, L. & Bustamante, C. Direct mechanical measurements of the elasticity of single DNA groups by using magnetic beads. *Science* **258**, 1122–1126 (1992).
15. Manning, G. S. A procedure for extracting persistence lengths from light-scattering data on intermediate molecular weight DNA. *Biopolymers* **20**, 1751–1755 (1981).
16. Smith, S. B., Yujia, C. & Bustamante, C. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA groups. *Science* **271**, 795–799 (1996).
17. Stryer, L. & Haugland, R. P. Energy transfer: a spectroscopic ruler. *Proc. Natl Acad. Sci. USA* **58**, 719–726 (1967).
18. Heller, M. J. & Morrison, L. E. in *Rapid Detection and Identification of Infectious Agents* (eds Kingsbury, D. T. & Falkow, S.) 245–256 (Academic, New York, 1985).
19. SantaLucia, J. Jr A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci. USA* **95**, 1460–1465 (1998).
20. Record, M. T. Jr, Anderson, C. F. & Lohman, T. M. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. *Q. Rev. Biophys.* **11**, 103–178 (1978).
21. Bockelmann, U., Essevaz-Roulet, B. & Heslot, F. Molecular stick-slip motion revealed by opening DNA with piconewton forces. *Phys. Rev. Lett.* **79**, 4489–4492 (1997).
22. Svoboda, K., Schmidt, C. F., Schnapp, B. J. & Block, S. M. Direct observation of kinesin stepping by optical trapping interferometry. *Nature* **365**, 721–727 (1993).
23. Kuo, S. C. & Sheetz, M. P. Force of single kinesin groups measured with optical tweezers. *Science* **260**, 232–234 (1993).
24. Finer, J. T., Simmons, R. M. & Spudis, J. A. Single myosin group mechanics: piconewton forces and nanometre steps. *Nature* **368**, 113–119 (1994).
25. Ishijima, A. *et al.* Single-group analysis of the actomyosin motor using nano-manipulation. *Biochem. Biophys. Res. Commun.* **199**, 1057–1063 (1994).
26. Green, C. & Tibbetts, C. Reassociation rate limited displacement of DNA strands by branch migration. *Nucleic Acids Res.* **9**, 1905–1918 (1981).
27. Lee, C. S., Davis, R. W. & Davidson, N. A physical study by electron microscopy of the terminally repetitive, circularly permuted DNA from the coliphage particles of *Escherichia coli* 15. *J. Mol. Biol.* **48**, 1–22 (1970).
28. Wetmur, J. G. & Davidson, N. Kinetics of renaturation of DNA. *J. Mol. Biol.* **31**, 349–370 (1968).
29. Radding, C. M., Beattie, K. L., Holloman, W. K. & Wiegand, R. C. Uptake of homologous single-stranded fragments by superhelical DNA: IV branch migration. *J. Mol. Biol.* **116**, 825–839 (1977).
30. Turberfield, A. J., Yurke, B. & Mills, A. P. Jr Coded self-assembly of DNA nanostructures. *Bull. Am. Phys. Soc.* **44**, 1711 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

F.C.S. thanks the Alexander von Humboldt Foundation for support.

Correspondence and requests for materials should be addressed to B.Y. (e-mail: yurke@lucent.com) or A.J.T. (e-mail: a.turberfield@physics.ox.ac.uk).

Influencing intramolecular motion with an alternating electric field

Veronica Bermudez*, Nathalie Capron†, Torsten Gase*,
Francesco G. Gatti‡, François Kajzar*, David A. Leigh‡,
Francesco Zerbetto† & Songwei Zhang‡

* LETI (CEA - Technologies Avancées), DEIN/SPE/GCO, CE Saclay,
91191 Gif-sur-Yvette Cedex, France

† Dipartimento di Chimica "G. Ciamician", Università degli Studi di Bologna,
V. F. Selmi 2, 40126, Bologna, Italy

‡ Centre for Supramolecular and Macromolecular Chemistry, Department of
Chemistry, University of Warwick, Coventry CV4 7AL, UK

Analogues of mechanical devices that operate on the molecular level^{1–5}, such as shuttles^{6–10}, brakes¹¹, ratchets^{12,13}, turnstiles¹⁴ and unidirectional spinning motors^{15,16}, are current targets of both synthetic chemistry and nanotechnology. These structures are designed to restrict the degrees of freedom of submolecular components such that they can only move with respect to each other in a predetermined manner, ideally under the influence of some external stimuli. Alternating-current (a.c.) electric fields are commonly used to probe electronic structure, but can also change the orientation of molecules^{17–19} (a phenomenon exploited in liquid crystal displays), or interact with large-scale molecular motions, such as the backbone fluctuations of semi-rigid polymers^{20,21}. Here we show that modest a.c. fields can be used to monitor and influence the relative motion within certain rotaxanes²², molecules comprising a ring that rotates around a linear 'thread' carrying bulky 'stoppers' at each end. We observe strong birefringence at frequencies that correspond to the rate at which the molecular ring pirouettes about the thread, with the frequency of maximum birefringence, and by inference also the rate of ring pirouetting giving rise to it, changing as the electric field strength is varied. Computer simulations and nuclear magnetic resonance spectroscopy show the ring rotation to be the only dynamic process occurring on a timescale corresponding to the frequency of maximum birefringence, thus confirming that mechanical motion within the rotaxanes can be addressed, and to some extent controlled, by oscillating electric fields.

During routine investigations of the influence of discrete large-amplitude internal molecular motions on optical properties, an a.c. electric field was used to address the structures of two hydrogen-bond-assembled rotaxanes, **1** and **2**, and their uninterlocked molecular fragments. The rotaxanes differ from each other only in the nature of the thread; **1** contains two nitrene hydrogen-bond acceptors, and **2** a fumaramide unit (Fig. 1). Both threads template the formation of the benzylic amide macrocycle about them to form [2]rotaxanes in yields of 70 and 97%, respectively²³.

The a.c. field effects on **1** and **2** in dioxane (the nonpolar, non-electrically-conducting solvent of choice for such experiments²⁴) were monitored by the quadratic dependence of the refractive index, that is, by recording the electro-optic Kerr effect. (Put simply, the light travelling through a sample undergoes a time-independent phase shift accompanied, in an a.c. field, by a time-dependent phase shift of twice that frequency.) Figure 2a shows the experimentally determined Kerr constant for the two rotaxanes as a function of frequency. A major response was somewhat unexpectedly discovered for both rotaxanes at values close to 50 Hz at 0.35 V μ m⁻¹—an unusual frequency which cannot easily be attributed to any intrinsic electronic phenomenon. Signals in this frequency range are absent for the corresponding uninterlocked components, that is, the threads or macrocycle alone, or the solvent itself, indicating that the response is intrinsically related to the mechanically interlocked architecture of the rotaxanes. Concentration-dependent measure-

ments show the expected linear dependence of the response. The generated signals are large; at rotaxane concentrations of $10^{-7} \text{ mol dm}^{-3}$ the observed Kerr constant is comparable to that of liquid nitrobenzene²⁵, and the frequency is dependent upon both the field strength, shifting to lower values with increasing voltage (Fig. 2b), and temperature, increasing at higher temperatures. Thus, whatever is causing the phenomenon is addressable by both a

tunable electric field and thermal stimulus. A second, higher-frequency response (around 80 Hz depending on field strength and temperature) was found for **2** but not for **1**.

What gives rise to these phenomena and why do the rotaxanes behave slightly differently one to another and very differently to their uninterlocked, but otherwise chemically identical, components? If the a.c. field were addressing an intermolecular alignment

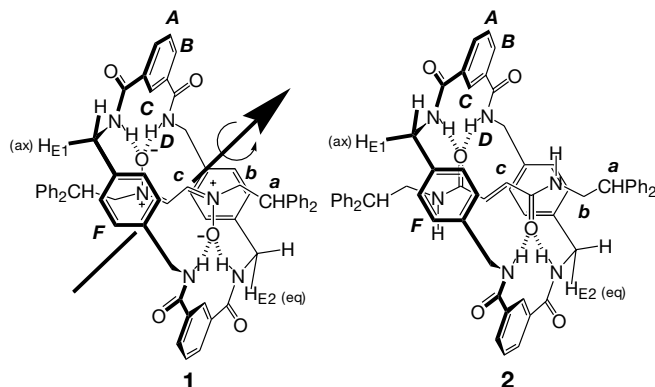


Figure 1 Rotaxanes **1** and **2** showing preferred hydrogen bonding motifs (from MM3 calculations, ^1H NMR spectroscopy, and X-ray crystallography of **2** and the bis-exopyridylmacrocycle analogue of **1**). Letters A, B, C, D, E1, E2, F, a, b and c identify non-

equivalent ^1H environments in each rotaxane. 180° rotation about the axis of the arrow, plus chair–chair flipping of the macrocycle, translates the $\text{H}_{\text{E}2}$ (equatorial) protons onto the $\text{H}_{\text{E}1}$ (axial) sites.

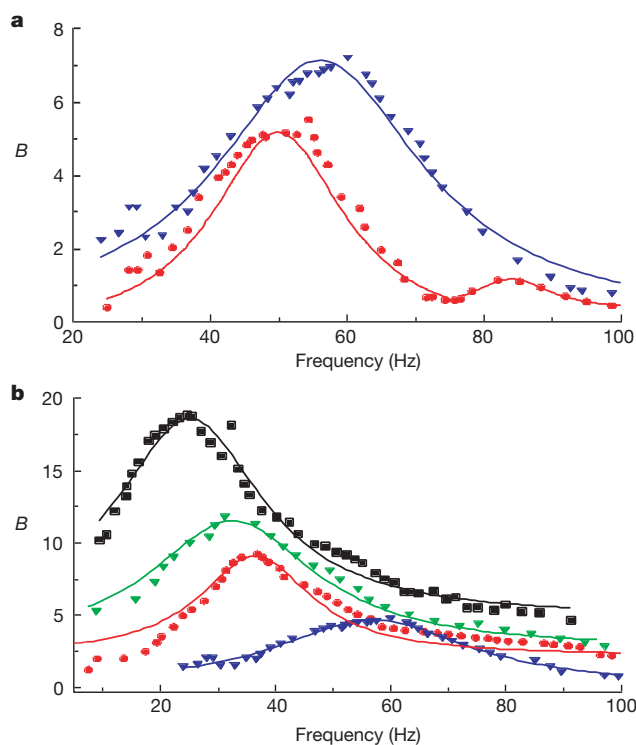


Figure 2 Kerr-effect measurements (B in pm V^{-2}) for rotaxanes **1** and **2**. **a**, Kerr constant of **1** (blue) and **2** (red) as a function of the a.c. frequency of the electric field in dioxane at 296 K. Applied voltage (U_{eff}) = 7 V; separation between electrodes = $20 \mu\text{m}$ (that is, an electric field strength of $0.35 \text{ V } \mu\text{m}^{-1}$). Solid lines are Lorentz fits (maxima at 56 Hz, **1**; and 50 and 84 Hz, **2**). **b**, Field-strength-dependence of the Kerr constant of **1** as a function of the a.c. frequency of the electric field at 296 K. Applied voltage (U_{eff}) = 7 V (blue, maxima at 56 Hz), 11 V (red, maxima at 36 Hz), 13 V (green, maxima at 33 Hz) and 16 V (black, maxima at 25 Hz). Extrapolation of a polynomial fit ($Y = A_0 + A_1X + A_2X^2 + A_3X^3 \dots$) of the experimental data to field strengths at which the Kerr effect is not measurable with our experimental set-up gives frequencies of 160 Hz at 4 V, 350 Hz at 2 V, 495 Hz at 1 V, 605 Hz at 0.4 V...and so on.

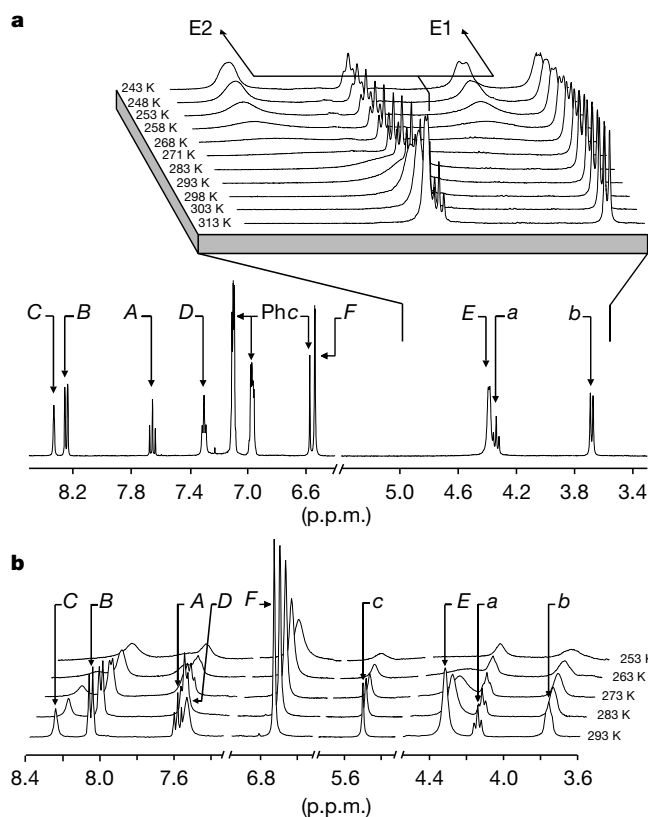


Figure 3 Determination of the dynamic processes occurring in rotaxanes **1** and **2** by variable temperature NMR experiments. **a**, **b**, 400 MHz ^1H NMR spectra (1:4 $\text{C}_4\text{D}_8\text{O}_2/\text{C}_2\text{D}_2\text{Cl}_4$) of **1**, full spectrum at 313 K, partial spectrum between 243 and 313 K shown in expansion (**a**), and **2** between 253 and 293 K (**b**). The letters correspond to the assignment of protons shown on the chemical structures of **1** and **2** in Fig. 1.

of dipoles, the timescale should be sub-nanosecond rather than tens of milliseconds per event (even water, a strongly intermolecularly hydrogen-bonded system, orients in the range 1.2 to 13 ps)^{17–19}. Instead, this time regime is generally the domain of large-scale intramolecular motions, a type of dynamics and timeframe which can sometimes be probed by nuclear magnetic resonance (NMR) experiments.

Accordingly, **1** and **2** were investigated by variable temperature (VT) ¹H NMR spectroscopy²⁶ (Fig. 3a and b) in a 1:4 mixture of dioxane-*d*₈ and 1,1,2,2-tetrachloroethane-*d*₂ (dioxane alone is not practical as a solvent for low-temperature studies as its melting point is 11 °C and tetrachloroethane is the least polar solvent in which the rotaxanes have good solubility over the required temperature range). The ¹H NMR spectrum of **1** at 313 K (Fig. 3a) shows the minimum number of signals possible, indicating the operational equivalency of each set of chemically distinct protons, including all eight macrocycle methylene protons (H_E). Cooling the sample to 243 K, however, leads to the resolution of the methylene protons into two non-interconverting magnetically distinct environments. No other proton resonances are split at this energy barrier (Fig. 3a, inset) indicating that the motion being frozen out prevents only an interconversion of non-equivalent H_E environments. The only process that satisfies this criterion is a 180° rotation of the macrocycle about the thread accompanied by a formal low-energy intracomponent chair–chair flip of the macrocycle). Coalescence²⁶ and spin polarization transfer by selective inversion recovery²⁷ (SPT-SIR) measurements give an energy barrier for this process of 12.2 ± 0.1 kcal mol^{–1} and a rate of macrocycle circumrotation (that is, two half turns) of 400 Hz at 271 K—a broadly similar time regime to the Kerr-effect response. Similar studies have been used to monitor various ring motions in related benzylic amide catenane and rotaxane structures that occur throughout the 10^{–6} to 10³ s^{–1} time-scale at ambient temperature^{23,28}. Consequently, the lack of changes in other signals in the VT NMR spectra indicates that macrocycle rotation is the only appreciable large-amplitude intramolecular or intercomponent motion occurring over a wide time frame.

Analogous VT NMR experiments (Fig. 3b) on rotaxane **2** gave a rate of macrocycle rotation of 340 Hz and a barrier of 11.4 ± 0.1 kcal mol^{–1} at 253 K. However, the temperature-dependent broadening and shifts of other protons in addition to H_E (Fig. 3b) show that other intramolecular motions are occurring on a similar timescale to circumrotation. It is interesting to recall that only a single signal (~50 Hz) is seen in the Kerr effect of **1** whereas two signals (~50 and ~80 Hz) are observed in this frequency

regime for **2**.

The dynamics of both rotaxane systems were simulated along lines previously reported²⁸ for related benzylic amide catenane systems using the MM3 potential²⁹ and the TINKER³⁰ program. The transition states, and atomic motion vectors connecting them to their minima, are shown in Fig. 4. As indicated by the NMR experiments, the simulations show that in **1** the macrocyclic ring rotates smoothly about the thread with little further perturbation. In comparison **2** presents a more complicated picture, where pirouetting is coupled to pivoting of the macrocycle against the thread, with the ideal fulcrum of motion located near the centre of mass of the molecule. The calculated activation energies, 11.2 kcal mol^{–1} for **1** and 13.9 kcal mol^{–1} for **2**, compare well with those determined by NMR.

In the absence of any other explanation, the peaks present in the Kerr effect of both rotaxanes must be a response to large-amplitude intramolecular motions, as previously observed with fluctuational motions of polymer backbones^{20,21}. Here, however, the NMR results, simulations, and the absence of the effect when the rotaxane components are not interlocked, indicate that the internal motion can be attributed to macrocycle pirouetting. In other words, birefringence is at a maximum when the a.c. electric field matches the macrocycle spinning rate.

A number of further points are worth emphasizing:

(1) The Kerr-constant frequency increases with temperature are mirrored by the increases in rates of circumrotation shown by NMR and the calculations.

(2) For solutions of **2**, a smaller peak exists at higher frequencies (for example, 84 Hz at 296 K, 0.35 V μm^{–1}), a feature consistent with the more complicated motion seen, but not resolved, by NMR. This suggests that the a.c. electric field may uncouple the two components of the composite motion revealed by the calculations.

(3) Differences in the rate constants found in the Kerr-effect experiments in dioxane and the NMR experiments in the dioxane/tetrachloroethane mixture are broadly in line with expectations. First, the rotaxane rings rotate more slowly in the less polar solvent because of stronger intercomponent hydrogen bonding. Second, because the birefringence peaks shift to lower frequencies at higher field strengths, the rings rotate faster during the NMR experiments where the applied electric field is zero. In fact, a simple polynomial correlation of the Kerr constant frequency maxima and the effective field strength suggests a room-temperature rate of rotation for the macrocycle in **1** of more than 600 Hz at fields of less than 0.4 V (Fig. 2b). □

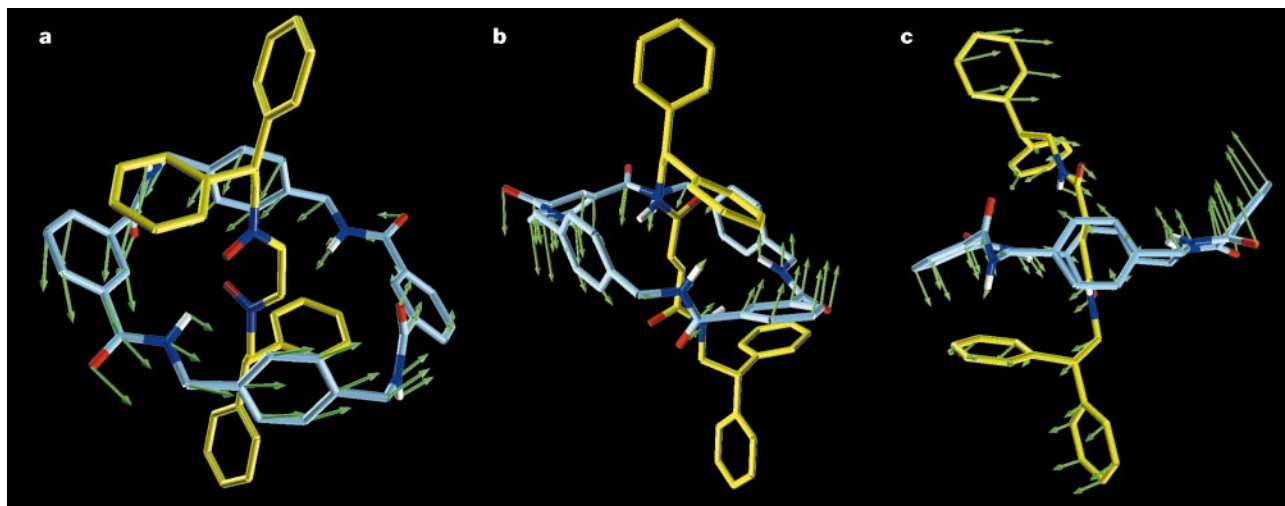


Figure 4 Calculated transition-state structures for macrocyclic ring motions for rotaxanes **1** and **2**. Green arrows represent the corresponding atomic motion vectors connecting the transition states to their minima. For clarity, only the vectors for the macrocycle atoms are

shown in **a** and **b**. **a**, Pirouetting of the macrocycle in **1**; **b**, pirouetting of the macrocycle in **2**; **c**, scissoring/pivoting motion in **2**.

Methods

Synthesis

Rotaxanes **1** and **2** were prepared from the respective threads, triethylamine, isophthaloyl dichloride and xylene diamine in chloroform by a procedure similar to that described previously for peptide-based threads²³. The detailed synthesis and characterization of these systems will be reported elsewhere.

Kerr-effect experiments

The birefringence measurements were carried out at dilute concentrations (10^{-7} mol dm⁻³) where rotaxane aggregation does not occur. Two parallel electrodes were placed in a cell between crossed polarizers with the polarizer axis and orthogonal analyser set perpendicularly to the electrodes. The passing light was gathered by a Si photodiode connected to a current-voltage converter that was protected by a monochromator to avoid stray light. The light source was a 5 mW He-Ne laser with a wavelength of 633 nm. The applied voltage was varied in the range 6–16 V, with a 20 μ m separation between electrodes consisting of two glass plates covered by a layer of indium tin oxide (ITO) about 50 nm thick. Similar apparatus has been previously described²⁴.

Received 23 February; accepted 17 June 2000.

1. Feynman, R. P. in *Miniaturization* (ed. Gilbert, H. D.) 282–296 (Reinhold, New York, 1961).
2. Mislow, K. Molecular machinery in organic chemistry. *Chemtracts – Org. Chem.* **2**, 151–174 (1989).
3. Balzani, V., Gómez-López, M. & Stoddart, J. F. Molecular machines. *Acc. Chem. Res.* **31**, 405–414 (1998).
4. Sauvage, J.-P. Transition metal-containing rotaxanes and catenanes in motion: toward molecular machines and motors. *Acc. Chem. Res.* **31**, 611–619 (1998).
5. Gimzewski, J. Molecules, nanophysics and nanoelectronics. *Phys. World* **11**, 29–33 (1998).
6. Anelli, P. -L., Spencer, N. & Stoddart, J. F. A molecular shuttle. *J. Am. Chem. Soc.* **113**, 5131–5133 (1991).
7. Bissell, R. A., Córdova, E., Kaifer, A. E. & Stoddart, J. F. A chemically and electrochemically switchable molecular shuttle. *Nature* **369**, 133–137 (1994).
8. Collin, J. -P., Gavina, P. & Sauvage, J.-P. Electrochemically induced molecular motions in copper-complexed threaded systems: From the unstoppered compound to the semi-rotaxane and the fully blocked rotaxane. *New J. Chem.* **21**, 525–528 (1996).
9. Lane, A. S., Leigh, D. A. & Murphy, A. Peptide-based molecular shuttles. *J. Am. Chem. Soc.* **119**, 11092–11093 (1997).
10. Murakami, H., Kawabuchi, A., Kotoo, K., Kunitake, M. & Nakashima, N. A light-driven molecular shuttle based on a rotaxane. *J. Am. Chem. Soc.* **119**, 7605–7606 (1997).
11. Kelly, T. R. *et al.* A molecular brake. *J. Am. Chem. Soc.* **116**, 3657–3658 (1994).
12. Kelly, T. R., Tellitu, I. & Sestelo, J. P. In search of molecular ratchets. *Angew. Chem. Int. Edn Engl.* **36**, 1866–1868 (1997).
13. Davis, A. P. Tilting at windmills? The second law survives. *Angew. Chem. Int. Edn Engl.* **37**, 909–910 (1998).
14. Bedard, T. C. & Moore, J. S. Design and synthesis of a “molecular turnstile”. *J. Am. Chem. Soc.* **117**, 10662–10671 (1995).
15. Kelly, T. R., De Silva, H. & Silva, R. A. Unidirectional rotary motion in a molecular system. *Nature* **401**, 150–152 (1999).
16. Koumura, N., Zijlstra, R. W. J., van Delden, R. A., Harada, N. & Feringa, B. L. Light-driven monodirectional molecular rotor. *Nature* **401**, 152–155 (1999).
17. Barthel, J., Bachhuber, K., Buchner, R. & Hetzenauer, H. Dielectric spectra of some common solvents in the microwave region - water and lower alcohols. *Chem. Phys. Lett.* **165**, 369–373 (1990).
18. Castner, E. W., Chang, Y. J., Chu, Y. C. & Walrafen, G. E. The intermolecular dynamics of liquid water. *J. Chem. Phys.* **102**, 653–659 (1995).
19. Cook, D. J., Chen, J. X., Morlino, E. A. & Hochstrasser, R. M. Terahertz-field-induced second-harmonic generation measurements of liquid dynamics. *Chem. Phys. Lett.* **309**, 221–228 (1999).
20. Kapitulnik, A., Casalnuovo, S., Lim, K. C. & Heeger, A. J. Electric-field coupling to slow elastic modes in gels of conjugated polymers. *Phys. Rev. Lett.* **53**, 469–472 (1984).
21. Lim, K. C., Kapitulnik, A., Zacher, R. & Heeger, A. J. Conformation of polydiacetylene macro-molecules in solution: Field-induced birefringence and rotational diffusion constant. *J. Chem. Phys.* **82**, 516–521 (1985).
22. Sauvage, J. -P. & Dietrich-Buchecker, C. O. (eds) *Molecular Catenanes, Rotaxanes and Knots* (Wiley-VCH, Weinheim, 1999).
23. Leigh, D. A., Murphy, A., Smart, J. P. & Slawin, A. M. Z. Glycylglycine rotaxanes – the hydrogen bond directed assembly of synthetic peptide rotaxanes. *Angew. Chem. Int. Edn Engl.* **36**, 728–732 (1997).
24. Kippelen, B., Sandalphon, Meerholz, K. & Peyghambarian, N. Birefringence, Pockels, and Kerr effects in photorefractive polymers. *Appl. Phys. Lett.* **68**, 1748–1750 (1996).
25. Buckingham, A. D. in *Molecular Electro-optics* (ed. O’Konski, C. T.) 7–62 (Marcel Dekker, New York, 1976).
26. Sandström, J. *Dynamic NMR Spectroscopy* (Academic, London, 1982).
27. Dahlquist, F. W., Longmur, K. J. & Du Vernet, R. B. Direct observation of chemical exchange by a selective pulse NMR technique. *J. Magn. Reson.* **17**, 406–410 (1975).
28. Leigh, D. A., Murphy, A., Smart, J. P., Deleuze, M. S. & Zerbetto, F. Controlling the frequency of macrocyclic ring rotation in benzylic amide [2]catenanes. *J. Am. Chem. Soc.* **120**, 6458–6467 (1998).
29. Allinger, N. L., Yuh, Y. H. & Li, J. -H. Molecular mechanics. The MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.* **111**, 8551–8582 (1989).
30. Ponder, J. W. & Richards, F. M. An efficient Newton-like method for molecular mechanics energy minimization of large molecules. *J. Comp. Chem.* **8**, 1016–1024 (1987).

Acknowledgements

We thank L. Joubert for his assistance in producing the visual representations of the rotaxane dynamics. This work was supported through the DRUM TMR network. F.Z.

acknowledges support from the MURST project “Dispositivi Supramolecolari”. D.A.L. is an EPSRC Advanced Research Fellow; F.G.G. is a Marie Curie Research Fellow. The Warwick group were responsible for the synthesis and NMR experiments, the Bologna group the simulations, and the Paris group the Kerr effect measurements.

Correspondence and requests for materials should be addressed to D.A.L. (e-mail: David.Leigh@Warwick.ac.uk) or F.Z. (e-mail: gatto@ciam.unibo.it).

Water activity as the determinant for homogeneous ice nucleation in aqueous solutions

Thomas Koop*, Beiping Luo†, Athanasios Tsias† & Thomas Peter*

* Swiss Federal Institute of Technology, Institute for Atmospheric Science, ETH Hönggerberg HPP, 8093 Zurich, Switzerland

† Max Planck Institute for Chemistry, Nachwuchsgruppe, Postfach 3060, 55020 Mainz, Germany

The unique properties of water in the supercooled (metastable) state are not fully understood¹. In particular, the effects of solutes and mechanical pressure on the kinetics of the liquid-to-solid phase transition of supercooled water and aqueous solutions to ice have remained unresolved. Here we show from experimental data that the homogeneous nucleation of ice from supercooled aqueous solutions is independent of the nature of the solute, but depends only on the water activity of the solution—that is, the ratio between the water vapour pressures of the solution and of pure water under the same conditions. In addition, we show that the presence of solutes and the application of pressure have a very similar effect on ice nucleation. We present a thermodynamic theory for homogeneous ice nucleation, which expresses the nucleation rate coefficient as a function of water activity and pressure. Recent observations from clouds containing ice are in good agreement with our theory and our results should help to overcome one of the main weaknesses of numerical models of the atmosphere, the formulation of cloud processes.

Solutes affect the equilibrium and non-equilibrium properties of water. The depression of the ice equilibrium melting point is one such example. In textbook derivations that apply to ideal solutions, the melting point depression is shown to depend only on a solute’s concentration (c), but is independent of its nature. However, at high concentrations large deviations from this approximation are observed (solid squares in Fig. 1a). On the other hand, the liquid-to-solid nucleation of ice from water and aqueous solutions is a kinetic non-equilibrium process. Ice nucleates homogeneously from micrometre-sized water droplets at about 235 K and the addition of solutes decreases this temperature even further (solid circles in Fig. 1a). The large variation in melting temperatures, T_m , and freezing temperatures, T_f , reflects the non-ideal behaviour of the solutions at moderate to high concentrations. To account for the non-ideality of the solutions we show in Fig. 1b the same data as a function of the concentration-, temperature- and pressure-dependent activity of water in solution, $a_w(c, T, p)$. (Here we refer to pure liquid bulk water as the reference state $a_w^0(0, T, p) \equiv a_w^0(T, p) = 1$.) We note that in equilibrium with water vapour a_w is equivalent to relative humidity, RH. In Fig. 1b the melting point curves of the various solutions collapse to a single line (dashed) because the a_w of a solution in equilibrium with ice, $a_w^i(T, p)$, is independent of the nature of the solute, that is, it is a function of only T and p (see equations (1) and (2) in Table 1). Also the freezing temperatures form a very compact distribution of data points, which is surprising

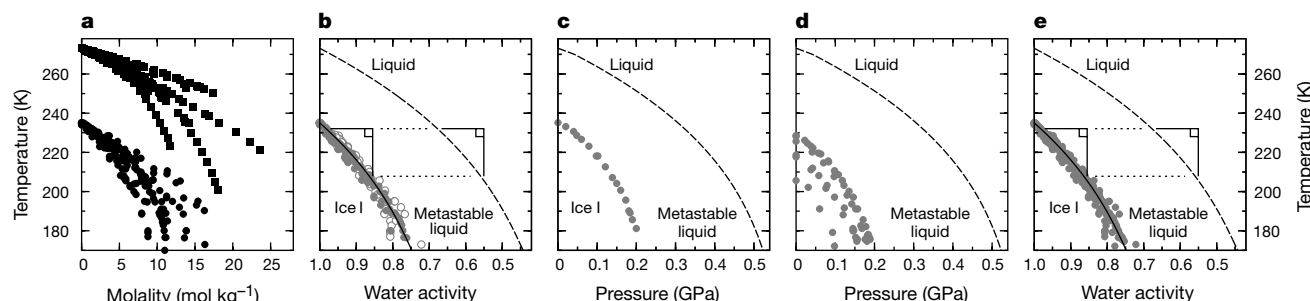


Figure 1 Experimental data used in the present analysis. **a**, Melting temperatures, T_m (squares), and freezing temperatures, T_f (circles), of droplets ($\sim 1\text{--}10\ \mu\text{m}$) of 18 different aqueous solutions as a function of the solute molality: H_2SO_4 , HNO_3 , ternary $\text{HNO}_3/\text{H}_2\text{SO}_4$, NH_4HSO_4 , $(\text{NH}_4)_2\text{SO}_4$, NH_4F , LiCl , NaCl , KCl , NH_4Cl , CaCl_2 , MnCl_2 , $\text{Ca}(\text{NO}_3)_2$, H_2O_2 , urea, ethylene glycol, glycerol and glucose^{5,6,8,9,18–24}. All salts and strong acids were treated to be fully dissociated and the molality is the resulting total ion molality. We have not used data obtained from Fourier transform infrared aerosol measurements because more recent studies employing emulsion and microscope techniques suggest that these data are subject to large uncertainties in composition^{18,21}. **b**, The same data as in **a**, as a function of the activity of water in solution, $a_w(c, T, O)$. For species shown as filled circles a_w

was calculated at T_f using an ion interaction model²⁵. For others (open circles) a_w was estimated by determining a_w at the melting temperature (dashed line), and assuming that for a fixed composition a_w is independent of temperature between T_m and T_f . The solid line is the melting point curve shifted by $\Delta a_w = 0.305$, as indicated by the two right angles connected by the dotted lines. **c**, Freezing temperatures of pure water droplets ($\sim 3.5\ \mu\text{m}$) as a function of pressure⁴. The dashed line is the melting point curve for ice I (ref. 3). **d**, Freezing temperatures of droplets of various aqueous NaCl and KCl solutions as a function of pressure⁶. **e**, All data points of **b–d** shown as $a_w(c^{\text{eff}}, T, O)$ calculated according to equation (6) in Table 1. The solid line is the same as in **b**.

given the large number of solutes. These included strongly dissociating acids and salts as well as weakly dissociating organics like glucose or ethylene glycol. In addition, the T_f curve is observed to be horizontally parallel to the T_m curve. This is indicated by the solid line which we constructed by adding a constant offset Δa_w to the a_w^i curve (dashed line) such that the solid line intersects the y -axis ($a_w = 1$ or RH = 1) at 235 K, the homogeneous freezing temperature of pure water in micrometre-sized droplets. The solid line describes the data points with a correlation coefficient of $r^2 = 0.96$ providing strong evidence that the kinetic (non-equilibrium) ice nucleation process is driven entirely by the thermodynamic (equilibrium) quantity $a_w - a_w^i$. We term this the ‘water-activity criterion’ for homogeneous ice nucleation.

In contrast, classical nucleation theory suggests that the nucleation process is driven only in part by $a_w(c, T, p)$ and $a_w^i(T, p)$ (namely, in terms of the ice saturation ratio $S^i = a_w/a_w^i$), but also depends on how the solute affects the ice/solution interface energy, σ , and the diffusion activation energy for a water molecule to cross the solution/ice interface, Δg_d (ref. 2). This apparent contradiction is resolved when σ and Δg_d are assumed to depend only on a_w but not on the nature of the solute. We can reconcile both theories (in numerical simulations not presented here) by using functions $\sigma(a_w)$ and $\Delta g_d(a_w)$ with physically plausible dependencies on a_w .

In Fig. 1c we show the T_m and T_f curves for pure water as a function of pressure^{3,4}. The very similar shape of the curves when compared to those in Fig. 1b suggests that solutes or applied pressure have a very similar effect on T_m and T_f . A similar connection between T_m and T_f for solutions and pure water under pressure has been suggested in earlier studies^{4–6}. What could be the explanation for these experimental observations? It has been shown by neutron diffraction experiments that the hydrogen bonding network is strongly affected by the presence of solutes or by pressure⁷. Large changes in the hydrogen–hydrogen (H–H) radial distribution function have been observed when comparing moderately concentrated NaCl solutions with pure water at 293 K. A similar change in the H–H radial distribution function was observed when applying high pressures to pure water. The neutron diffraction experiments⁷ indicate that the addition of 4 mol NaCl per kg of water has an effect on the water structure that is equivalent to increasing the pressure applied to pure water to 0.14 GPa. We compared T_m and T_f for these two cases and find $T_m = 258\ \text{K}$ and $T_f = 208\ \text{K}$ for a 4 mol kg^{-1} NaCl solution^{6,8,9} and $T_m = 259\ \text{K}$ and $T_f = 207\ \text{K}$ for pure water at 0.14 GPa (refs 3, 4). The very good agreement between these T_m and T_f values suggests that solutions and pure water under pressure melt and freeze when their hydrogen bonding network is the same.

Table 1 Parameterization of the water-activity criterion for the homogeneous ice nucleation rate coefficient J

Equation	Validity range	References
1 $a_w(c, T, p) = \exp\{[\mu_w(c, T, p) - \mu_w^0(T, p)]/RT\}$		
2 $\mu_w(T, 0) - \mu_w^0(T, 0) = 210368 + 131.438T - 3.32373 \times 10^6 T^{-1} - 41729.1 \ln(T)$	$150 < T < 273\ \text{K}$	12
3 $\int [\mu_w(c, T, p) - \mu_w^0(T, p)] dp \approx +v_w^0(T, 0)[p - 1/2\kappa_T^0(T, p)p^2 - 1/6[\partial\kappa_T^0(T, p)/\partial p]p^3] - v^i(T, 0)[p - 1/2\kappa_T^i(T, p)p^2 - 1/6[\partial\kappa_T^i(T, p)/\partial p]p^3]$	$0 < p < p_{\text{max}}(T)$	13, 14*
4 $v_w^0(T, 0) = -230.76 - 0.1478T + 4099.2T^{-1} + 48.8341 \ln(T)$	$170 < T < 240\ \text{K}$	15, 16
5 $v^i(T, 0) = 19.43 - 2.2 \times 10^{-3}T + 1.08 \times 10^{-5}T^2$	$150 < T < 273\ \text{K}$	2
6 $\Delta a_w(c, T, p) = a_w(c^{\text{eff}}, T, 0) - a_w^i(T, 0) = a_w(c, T, 0)\exp\{\int [\mu_w(c, T, p) - \mu_w^0(T, p)] dp/RT\} - a_w^i(T, 0)$	$170 < T < 240\ \text{K}$, $p < p_{\text{max}}(T)$	
7 $\log(J) = -906.7 + 8502\Delta a_w - 26924(\Delta a_w)^2 + 29180(\Delta a_w)^3$	$0.26 < \Delta a_w < 0.34$	2

$a_w(c, T, p)$ is the activity of water in a solution in equilibrium with ice, T is temperature in K, p is the total mechanical pressure in GPa, and R is the ideal gas constant. $\mu_w(c, T, p)$ and $\mu_w^0(T, p) = \mu_w(0, T, p)$ are the temperature and pressure dependent chemical potentials of water in pure ice and pure liquid water, respectively, in J mol^{-1} . The parameterization in equation (2) represents the thermodynamically consistent function at ambient pressure ($p = 0$) presented as curve 3 in Fig. 6 of ref. 12. This is in agreement with the experimentally determined value at 150 K (ref. 17). The water activity of a solution under pressure, $a_w(c, T, p)$, can be expressed through the chemical potential of water in the same solution at zero pressure, $\mu_w(c, T, 0)$, by $RT \ln(a_w(c, T, p)) = \mu_w(c, T, p) - (\mu_w^0(T, p) = \mu_w(c, T, 0) + \int \mu_w(c, T, p) dp - \mu_w^0(T, 0) - \int \mu_w^0(T, p) dp)$. Here, $\mu_w(c, T, p)$ is the partial molar volume of water in the solution in $\text{cm}^3 \text{mol}^{-1}$, $v_w^0(T, 0) = v_w(0, T, 0)$ is the molar volume of pure liquid water and $v_w^i(T, 0)$ its value at ambient pressure, which we smoothly interpolated between measured water molar volumes ($\geq 235\ \text{K}$, ref. 15) and the molar volume of low density amorphous ice (77 K, ref. 16) in accordance with the approach used in ref. 1 for the construction of a thermodynamically plausible equation of state for liquid water. $\kappa_T^0(T, 0) = 1.6\ \text{GPa}^{-1}$ is the isothermal compressibility of pure water at ambient pressure and $\partial\kappa_T^0(T, 0)/\partial p = -8.8\ \text{GPa}^{-2}$ is its pressure dependence. Both values have been determined from fitting the T_f points of pure water as a function of pressure (Fig. 1c) to give the same a_w values as the solid line representing the solution data at ambient pressure (Fig. 1b) using equation (6). $v^i(T, 0)$ is the molar volume of hexagonal ice at ambient pressure, $\kappa_T^i(T, 0) = 0.22\ \text{GPa}^{-1}$ is the isothermal compressibility of hexagonal ice at ambient pressure, and $\partial\kappa_T^i(T, 0)/\partial p = -0.17\ \text{GPa}^{-2}$ is its pressure dependence^{13,14}. J is the homogeneous ice nucleation rate coefficient in $\text{cm}^{-3} \text{s}^{-1}$. $p_{\text{max}}(T)$ is temperature dependent with $p_{\text{max}}(T) \approx -0.93 + 1.37 \times 10^{-2}T - 4.12 \times 10^{-5}T^2$ in GPa.

* Equation (3) contains the approximation that $\kappa_T^0(T, p)$ and $\kappa_T^i(T, 0)$ are temperature-independent¹⁴ and that $v_w(c, T, p) = v_w^0(T, p)$.

We propose that a solution with concentration c under pressure p freezes at the same temperature as another solution with concentration c^{eff} under zero pressure, when the corresponding hydrogen bonding networks change in the same way upon freezing and the accompanying amounts of released energy are the same. This corresponds to a change in chemical potential of water $\mu_w(c, T, p) - \mu_w^i(T, p) = \mu_w(c^{\text{eff}}, T, 0) - \mu_w^i(T, 0)$, which serves to generalize the relation $\Delta a_w = a_w - a_w^i$ to the pressure-dependent results given in equation (6) of Table 1.

To study the simultaneous effects of solute and pressure on T_f and as a further test for the validity of our approach, we have used experimental data⁶ of pressurized aqueous NaCl and KCl solution droplets shown in Fig. 1d. We use equation (6) to transform the pressure-dependent data in Fig. 1c and d into effective-concentration-dependent (c^{eff}) data in order to compare all data of Fig. 1b, c and d in Fig. 1e. The very good agreement between all data points indicates that the homogeneous nucleation of ice is only a function of Δa_w .

We extend the water-activity criterion for homogeneous ice nucleation to determine nucleation rates. As homogeneous nucleation is a stochastic process, the nucleation rate, ω , is proportional to the volume, V , of the metastable solution, $\omega = JV$, where J is the homogeneous nucleation rate coefficient. Larger samples have higher homogeneous freezing temperatures than smaller samples. The T_f points shown in Fig. 1 were determined from droplets about 1–10 μm in diameter. Taking into account the experimental cooling rates, these T_f values correspond to a homogeneous nucleation rate coefficient of $J \approx 10^{10} \text{ cm}^{-3} \text{ s}^{-1}$. As shown above, this value of J depends only on Δa_w . Assuming this to be the case also for other values of J , we can compute $J(\Delta a_w)$ by fitting to the tabulated J -values of pure water² ($J(a_w = 1)$). The results are shown in Fig. 2a for selected values of J . A parameterization for $J(\Delta a_w)$ is given in equation (7) of Table 1. Figure 2a indicates that the derivative of J with respect to temperature, $\partial J / \partial T$, is larger the higher the temperature. In Fig. 2b we show as a solid line the slope $\partial[\log(J)] / \partial T|_{a_w = \text{const}}$ calculated from Table 1 for $J = 10^8 \text{ cm}^{-3} \text{ s}^{-1}$, which allows a direct comparison to experimentally determined $\partial J / \partial T$ values. The good agreement further supports our proposal of J being a function of Δa_w only.

Potential application of the water-activity criterion range from atmospheric ice cloud formation to geological processes like frost heave or biological ones like cell damaging due to ice freezing. In the following we restrict ourselves to ice clouds, which strongly affect both the chemistry and the radiative properties of the Earth¹⁰.

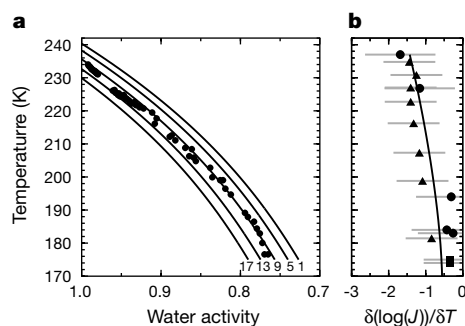


Figure 2 Variation of the homogeneous ice nucleation rate coefficient J with water activity and temperature. **a**, Isolines for selected values of J as a function of a_w and T . The numbers indicate the exponent, x , for $J = 10^x \text{ cm}^{-3} \text{ s}^{-1}$. Data points are the experimentally observed T_f values for aqueous H_2SO_4 (ref. 18). **b**, The line gives the slope $\partial[\log(J)] / \partial T|_{a_w = \text{const}}$ as a function of temperature for $J = 10^8 \text{ cm}^{-3} \text{ s}^{-1}$ as calculated from Table 1. Also shown are experimentally determined values for H_2SO_4 (circles)²⁶, NH_4HSO_4 (triangles, determined from the normalized crystallization peak width in emulsion experiments)²⁰, and LiCl (squares)²⁷ with their uncertainty shown as grey horizontal lines.

Despite their importance the microphysical mechanisms responsible for the formation of ice clouds are not very well understood and are one of the uncertainties in assessing the global radiation budget and future climate change¹⁰. Ice particles in the atmosphere can form through homogeneous ice nucleation in pre-existing liquid aerosols. These aerosols are in the submicrometre size range and are believed to consist predominantly of varying mixtures of H_2SO_4 , HNO_3 , NH_3 , and H_2O although a wide range of other species including organics have been observed as well¹¹. Using $J(\Delta a_w)$ we may compute the formation of atmospheric ice cloud particles through homogeneous nucleation as a function of temperature, aerosol size and water activity without knowing the composition of the aerosols. As mentioned above, water activity under atmospheric conditions may be equivalently expressed in terms of relative humidity, or for upper tropospheric/lower stratospheric purposes, more practically in terms of the ice saturation ratio S^i . Here we calculate S^i as a function of temperature and liquid aerosol size by choosing a rate $\omega = JV = 1 \text{ min}^{-1}$, which is equal to a freezing probability of $P = 1 - \exp(-JVt) = 0.63$ within a time $t = 1 \text{ min}$ (that is, about two-thirds of the aerosols would be frozen after 1 min). The results are shown in Fig. 3 as solid lines with the size indicated in the top left corner. We note that the Kelvin effect has not been considered in these calculations because its effect on the water activity through an increase in the internal mechanical pressure is of minor importance for droplets with $r \geq 0.05 \mu\text{m}$. We also show observations of atmospheric ice clouds at altitudes ranging from 8 to 22 km in Fig. 3. These are in very good agreement with our predicted behaviour of S^i as a function of temperature and aerosol size; thus, we believe these clouds were indeed induced by homogeneous ice nucleation from pre-existing liquid aerosols. The water-

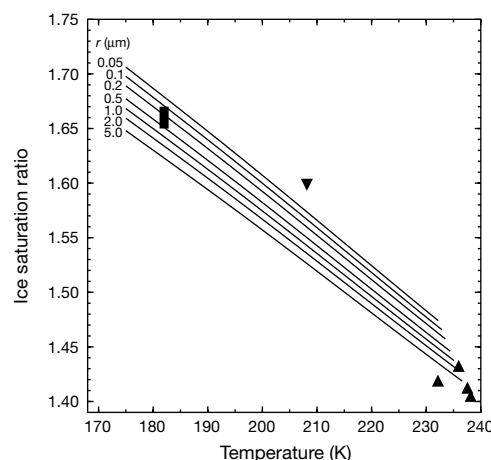


Figure 3 Ice saturation ratios, S^i , for different aerosol radii, r , as a function of temperature. Upward triangles indicate observations of liquid water droplets with the lidar technique (droplets are irradiated with a laser pulse and the backscattered light is detected) and *in situ* measurements of S^i in cirrus clouds at about 8 km altitude^{28,29} (the measured S^i values represent only lower limits to S^i required to nucleate ice). Downward triangles indicate an upper limit for S^i at 208 K from an *in situ* measurement of an orographically induced cirrus ice cloud at about 12 km altitude²⁹. Upward and downward triangles bracket our calculated lines for aerosols with radii between 0.05 and 5 μm . Rectangles indicate lidar observations of a polar stratospheric ice cloud at an altitude of about 22 km at 182 K evaluated by means of a mesoscale microphysical box model together with light scattering calculations to simulate the observed lidar images along the quasi-lagrangian trajectory of the cloud particles³⁰. (Instead of adjusting the ice particle number density in order to bring the calculated optical properties into agreement with the observations as in ref. 30, we used the homogeneous nucleation rates presented in Table 1 to initiate the nucleation and subsequent growth of the ice particles in the model; the only free parameter in these calculations is the unknown absolute trajectory temperature which we chose to be 1.1 K higher than in the original study, bringing the calculated and observed total and perpendicularly polarized backscattering ratios into good agreement).

activity criterion is hence the basis for a thermodynamically consistent new parameterization for J that improves and facilitates the simulation of homogeneous nucleation of ice clouds. Important applications range from issues like denitrification and future ozone depletion in the Arctic polar stratosphere to the aerosol indirect effect owing to homogeneously formed ice clouds in climate studies. The simplicity of the formulation makes it suitable not only for process studies but also for global-scale three-dimensional models. □

Received 3 December 1999; accepted 16 June 2000.

- Mishima, O. & Stanley, H. E. The relationship between liquid, supercooled and glassy water. *Nature* **396**, 329–335 (1998).
- Pruppacher, H. R. & Klett, J. D. *Microphysics Of Clouds And Precipitation* 79–80 & 205–215 (Kluwer, Dordrecht, 1997).
- Mishima, O. Relationship between melting and amorphization of ice. *Nature* **384**, 546–549 (1996).
- Kanno, H., Speedy, R. & Angell, C. A. Supercooling of water to -92°C under pressure. *Science* **189**, 880–881 (1975).
- Rasmussen, D. H. Thermodynamics and nucleation phenomena - a set of experimental observations. *J. Cryst. Growth* **56**, 56–66 (1982).
- Kanno, H. & Angell, C. A. Homogeneous nucleation and glass formation in aqueous alkali halide solutions at high pressures. *J. Phys. Chem.* **81**, 2639–2643 (1977).
- Leberman, R. & Soper, A. K. Effect of high salt concentrations on water structure. *Nature* **378**, 364–366 (1995).
- Rasmussen, D. H. & MacKenzie, A. P. in *Water Structure at the Water Polymer Interface* (ed. Jellinek, H. G.) 126–145 (Plenum, New York, 1972).
- Weast, R. C. (ed.) *Handbook Of Chemistry And Physics* D139–D178 (Chemical Rubber Company, Cleveland, 1966).
- Baker, M. B. Cloud microphysics and climate. *Science* **276**, 1072–1078 (1997).
- Murphy, D. M., Thomson, D. S. & Mahoney, M. J. In situ measurements of organics, meteoritic material, mercury, and other elements in aerosols at 5 to 19 kilometers. *Science* **282**, 1664–1669 (1998).
- Johari, G. P., Fleissner, G., Hallbrucker, A. & Mayer, E. Thermodynamic continuity between glassy and normal water. *J. Phys. Chem.* **98**, 4719–4725 (1994).
- Gagnon, R. E., Kieffe, H., Cloutier, M. J. & Whalley, E. Pressure dependence of the elastic constants of ice Ih to 2.8 kbar by Brillouin spectroscopy. *J. Chem. Phys.* **89**, 4522–4528 (1988).
- Dantl, G. in *Physics of Ice* (eds Riehl, N., Bullemer, B. & Engelhardt, H.) 223–230 (Plenum, New York, 1969).
- Hare, D. E. & Sorensen, C. M. The density of supercooled water. II. Bulk samples cooled to the homogeneous nucleation limit. *J. Chem. Phys.* **87**, 4840–4845 (1987).
- Mishima, O., Calvert, L. D. & Whalley, E. An apparently first-order transition between two amorphous phases of ice induced by pressure. *Nature* **314**, 76–78 (1985).
- Speedy, R. J., Debenedetti, P. G., Smith, R. S., Huang, C. & Kay, B. D. The evaporation rate, free energy, and entropy of amorphous water at 150K. *J. Chem. Phys.* **105**, 240–244 (1996).
- Koop, T., Ng, H. P., Molina, L. T. & Molina, M. J. A new optical technique to study aerosol phase transitions: the nucleation of ice from H_2SO_4 aerosols. *J. Phys. Chem. A* **102**, 8924–8931 (1998).
- Chang, H. Y. A., Koop, T., Molina, L. T. & Molina, M. J. Phase transitions in emulsified $\text{HNO}_3/\text{H}_2\text{O}$ and $\text{HNO}_3/\text{H}_2\text{SO}_4/\text{H}_2\text{O}$ solutions. *J. Phys. Chem. A* **103**, 2673–2679 (1999).
- Koop, T., Bertram, A. K., Molina, L. T. & Molina, M. J. Phase transitions in aqueous NH_4HSO_4 solutions. *J. Phys. Chem. A* **103**, 9042–9048 (1999).
- Bertram, A. K., Koop, T., Molina, L. T. & Molina, M. J. Ice formation in $(\text{NH}_4)_2\text{SO}_4\text{-H}_2\text{O}$ particles. *J. Phys. Chem. A* **104**, 584–588 (2000).
- Angell, C. A., Sare, E. J., Donnelly, J. & MacFarlane, D. R. Homogeneous nucleation and glass transition temperatures in solutions of Li salts in D_2O and H_2O . Doubly unstable glass regions. *J. Phys. Chem.* **85**, 1461–1464 (1981).
- Oguni, M. & Angell, C. A. Heat capacities of $\text{H}_2\text{O}+\text{H}_2\text{O}_2$, and $\text{H}_2\text{O}+\text{N}_2\text{H}_4$ binary solutions: isolation of a singular component for c_p of supercooled water. *J. Chem. Phys.* **73**, 1948–1954 (1980).
- Schäfer, K. & Lax, E. (eds) *Landolt-Börnstein, Zahlenwerte und Funktionen* Part 2b (Springer, Berlin, 1962).
- Clegg, S. L., Brimblecombe, P. & Wexler, A. S. Thermodynamic model of the system $\text{H}^+ - \text{NH}_4^+ - \text{SO}_4^{2-} - \text{NO}_3^- - \text{H}_2\text{O}$ at tropospheric temperatures. *J. Phys. Chem. A* **102**, 2137–2154 (1998).
- Krämer, B. *Laboruntersuchungen zum Gefrierprozess in polaren stratosphärischen Wolken* PhD thesis, Free University Berlin (1998).
- MacFarlane, D. R., Kadiyala, R. K. & Angell, C. A. Homogeneous nucleation and growth of ice from solutions. TTT curves, the nucleation rate, and the stable glass criterion. *J. Chem. Phys.* **79**, 3921–3927 (1983).
- Sassen, K. & Dodd, G. C. Homogeneous nucleation rate for highly supercooled cirrus cloud droplets. *J. Atmos. Sci.* **45**, 1357–1369 (1988).
- Heymsfield, A. J., Miloshevich, L. M., Twohy, C., Sachse, G. & Oltmans, S. Upper-tropospheric relative humidity observations and implications for cirrus ice nucleation. *Geophys. Res. Lett.* **25**, 1343–1346 (1998).
- Carlsaw, K. S. *et al.* Particle microphysics and chemistry in remotely observed mountain polar stratospheric clouds. *J. Geophys. Res.* **103**, 5785–5796 (1998).

Acknowledgements

We are grateful to A. Bertram, C. Jeffery, G. Johari, O. Mishima, and H. Vortisch for helpful discussions and for providing us with original data sets. We also thank M. Canagaratna and J. Staehelin for helpful comments on manuscript drafts.

Correspondence and requests for materials should be addressed to T.K. (e-mail: koop@atmos.umnw.ethz.ch).

Evidence from three-dimensional seismic reflectivity images for enhanced melt supply beneath mid-ocean-ridge discontinuities

G. M. Kent*, S. C. Singh†‡§, A. J. Harding*, M. C. Sinha†§, J. A. Orcutt*, P. J. Barton§, R. S. White§, S. Bazin*‡, R. W. Hobbs†, C. H. Tong§ & J. W. Pye§

* Cecil H. and Ida M. Green Institute of Geophysics and Planetary Physics, University of California, San Diego, La Jolla, California 92093, USA

† British Institutions Reflection Profiling Syndicate (BIRPS), § Department of Earth Sciences, Bullard Laboratories, University of Cambridge, Cambridge CB3 0EZ, UK

Quantifying the melt distribution and crustal structure across ridge-axis discontinuities is essential for understanding the relationship between magmatic, tectonic and petrologic segmentation of mid-ocean-ridge spreading centres. The geometry and continuity of magma bodies beneath features such as overlapping spreading centres can strongly influence the composition of erupted lavas¹ and may give insight into the underlying pattern of mantle flow. Here we present three-dimensional images of seismic reflectivity beneath a mid-ocean ridge to investigate the nature of melt distribution across a ridge-axis discontinuity. Reflectivity slices through the $9^{\circ}03' \text{ N}$ overlapping spreading centre on East Pacific Rise suggest that it has a robust magma supply, with melt bodies underlying both limbs and ponding of melt beneath large areas of the overlap basin. The geometry of melt distribution beneath this offset is inconsistent with large-scale, crustal redistribution of melt away from centres of upwelling^{2,3}. The complex distribution of melt seems instead to be caused by a combination of vertical melt transport from the underlying mantle and subsequent focusing of melt beneath a magma freezing boundary in the mid-crust.

High-resolution swath mapping systems have provided dramatic images of mid-ocean plate boundaries, revealing a complex system that is segmented at a variety of scales^{2,4}. For fast spreading rates, large-scale segmentation is defined by widely spaced transform faults, whose offsets are large enough to accommodate rigid slip. Ridge segments between transforms are further divided by a smaller ($< 10 \text{ km}$) series of non-rigid plate boundaries, which can give the spreading centre a sinuous appearance. The largest non-transform offsets are recognized by their inward-curving limbs, which encircle an elongated depression^{5,6}, and are usually sited at deeper portions of the ridge crest. Some investigators have attributed the morphological character of these “overlapping spreading centres” to a low magma supply^{2,5}. These models of magmatic segmentation propose enhanced upwelling beneath the shallowest portions of the ridge crest, and, by inference, place large non-transform offsets along magmatically starved sections of the ridge, although there are models where along-axis flow is so efficient that magma supply is enhanced near ridge-axis discontinuities³. Petrologic sampling along mid-ocean ridges reveals a similar pattern of geochemical segmentation, suggesting an intimate connection between tectonic and magmatic segmentation of the ridge crest⁷. Other indicators of magmatic budget such as axial volume⁸ and mantle Bouguer anomaly⁹ tend to support the notion that the shallower, more

‡ Present addresses: Laboratoire de Sismologie, Institut de Physique du Globe de Paris, 75252 Paris Cedex 05, France (S.C.S., S.B.); School of Ocean and Earth Science, University of Southampton, Southampton SO14 3ZH, UK (M.C.S.).

inflated portions of the ridges are sites of enhanced magma injection from the mantle, which serve as the locus for redistribution of melt towards the distal ends of the segments. In contrast, seismic observations^{10–12} across smaller discontinuities reveal a crustal magma chamber that is segmented at a scale of 10–15 km, providing evidence for only small-scale mixing of magma along the axis. Furthermore, the handful of reflection profiles^{13–15} that cross overlapping spreading centres do not show a muted magma supply, but instead appear to show an abundance of melt near these features, suggesting a complex relationship between ridge-axis discontinuities and the underlying pattern of magmatic segmentation.

To differentiate between competing models of magmatic segmentation^{2,3,6}, we conducted a three-dimensional seismic reflection survey across the 9° 03' N overlapping spreading centre, East Pacific Rise (Fig. 1). The ARAD (anatomy of a ridge-axis discontinuity) seismic experiment took place on board the RV *Maurice Ewing* in 1997. Thirty on-bottom hydrophones were also deployed to define the crustal velocity structure. Coincident three-dimensional images of seismic reflectivity and velocity represent the best approach for understanding the detailed structure across this ridge-axis discontinuity. The spatially dense layout of reflection profiles and on-bottom receivers was a necessity given the three-dimensional nature of the target. In this setting, widely spaced two-

dimensional seismic profiles would provide insufficient sampling of out-of-plane wave propagation, resulting in significant imaging errors. In addition, a complex, highly reflective sea floor can scatter out-of-plane energy onto a two-dimensional profile, producing coherent artefacts within the section, which can be especially acute in the mid-ocean-ridge environment. To overcome these difficulties, a series of 201 reflection profiles separated by 100 m were collected during the ARAD experiment to allow a full three-dimensional processing of the data. This scheme ensures that virtually all of the seismic wavefield will be accurately stacked and migrated, generating detailed images of crustal structure across this ridge-axis discontinuity.

Reflectivity slices through the three-dimensional volume reveal continuous melt sill events beneath both limbs of the overlapping spreading centre, with some melt-based reflections found underlying the northern overlap basin. The geometry of the melt sill underlying the western limb is relatively simple, showing a modest narrowing in width, but little or no depth variation along the ridge crest before dying out (Figs 2 and 3). In contrast, the melt sill beneath the propagating¹⁶ eastern limb deepens significantly over its southernmost 6 km, plunging some 500 m in depth relative to the sea floor (Fig. 3). Unlike the melt sill beneath the western limb, which shows little morphological change with latitude, the melt sill

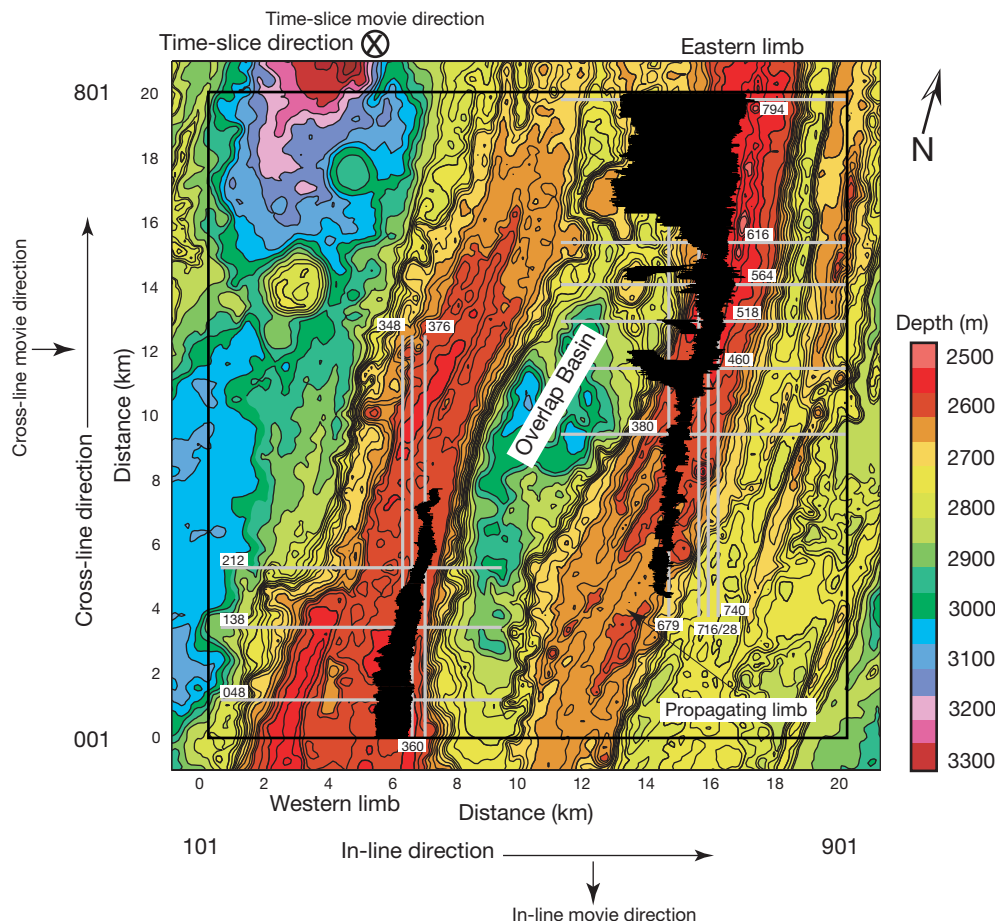


Figure 1 Sea-floor bathymetry and nomenclature for the ARAD seismic experiment. Sea-floor depths were obtained from redundant (~30-fold) multibeam soundings obtained during the acquisition of the three-dimensional seismic volume. Bathymetry is contoured at 50-m intervals, with shallow regions of the discontinuity coloured red/orange to highlight the overlapping nature of the offset. Thin white lines show the location of seismic profiles presented in Figs 2 and 3. The areal extent of crustal magma chambers are shown as black in-fill overlying sea-floor bathymetry. Seismic volume coordinates are also

indicated. The in-line direction corresponds to the sailing direction during data acquisition, while the cross-line direction is perpendicular to the sail-line azimuth. Individual in- and cross-line slices are separated by 25 m. In-line slices range from 1 to 801, while cross-line slices range from 101 to 901. Time-slices are oriented orthogonal to the plane defined by in- and cross-line directions, and are separated by 4-ms increments. Forward-play directions of Supplementary Information movie frames are also indicated.

geometry beneath the eastern limb is very complex. To the north, the melt horizon is displaced westward of the ridge-axis toward the overlap basin, and is some 4 km in width (Fig. 2, slice 794). The melt sill in this region also appears to be offset by a constant travel-time delay from the layer 2A reflection; if the layer 2A reflection represents the extrusive-dyke contact^{17–20}, then this geometry suggests that the melt sill may follow a permeable pathway just below the base of the sheeted-dyke complex. Ultimately, a ridge-centred melt lens²¹ becomes dominant along the southern reaches of the propagating ridge tip, but is occasionally supplied by discrete conduits of melt, which originate beneath the central overlap basin. This observation suggests that melt ascending beneath the overlap basin is an important source for lavas erupted along the propagating limb. It should also be noted that although the plunging melt sill beneath the propagating tip is continuous, its width remains remarkably narrow, typically less than 500 m over its last 3–4 km of existence.

Three-dimensional images of seismic reflectivity across this large non-transform offset do not show a continuous melt sill underlying the entire basin, which presumably rules out mixing of melt between the western and eastern limbs, although these ridge-crest magma reservoirs may have a common source beneath the overlap basin. The reflection images are further supported by on-bottom-recorded, wide-angle refracted arrivals, which clearly show a sig-

nificant low-velocity zone underlying the entire basin²². The near-axis permeability structure of the mid- to upper-crust is probably controlled by the base of the sheeted-dyke complex, plus any small amounts of isotropic gabbros which have solidified²³. Thus, melt ascending vertically beneath the basin floor will probably be focused along this permeability front, which may explain both the extended melt horizons found dipping upwards toward the neovolcanic zone along the eastern limb, and why these features, to first-order, are found at a constant travel-time beneath the presumed extrusive-dyke contact¹⁵. Moreover, regardless of magma budget, the likelihood of maintaining a continuous melt sill beneath the basin is low, as melt would have to coalesce beneath a permeability front, which is synclinal in form. Instead, it is more likely that the buoyant melt would migrate up-dip towards the neovolcanic zones beneath each limb, where there is closure within the permeability structure; this capture mechanism is analogous to the migration of buoyant gas and oil towards anticlinal structures where these fluids are then trapped against an impermeable barrier. Ultimately, the thermal state of the ridge crest²⁴ is controlled by the permeability structure, which sets the sheeted-dyke thickness, and the growth rate of isotropic gabbros, which together form a seal that focuses melt towards delivery along the neovolcanic zone.

A clear understanding of the three-dimensional structure beneath a large ridge-axis discontinuity provides new insights into the

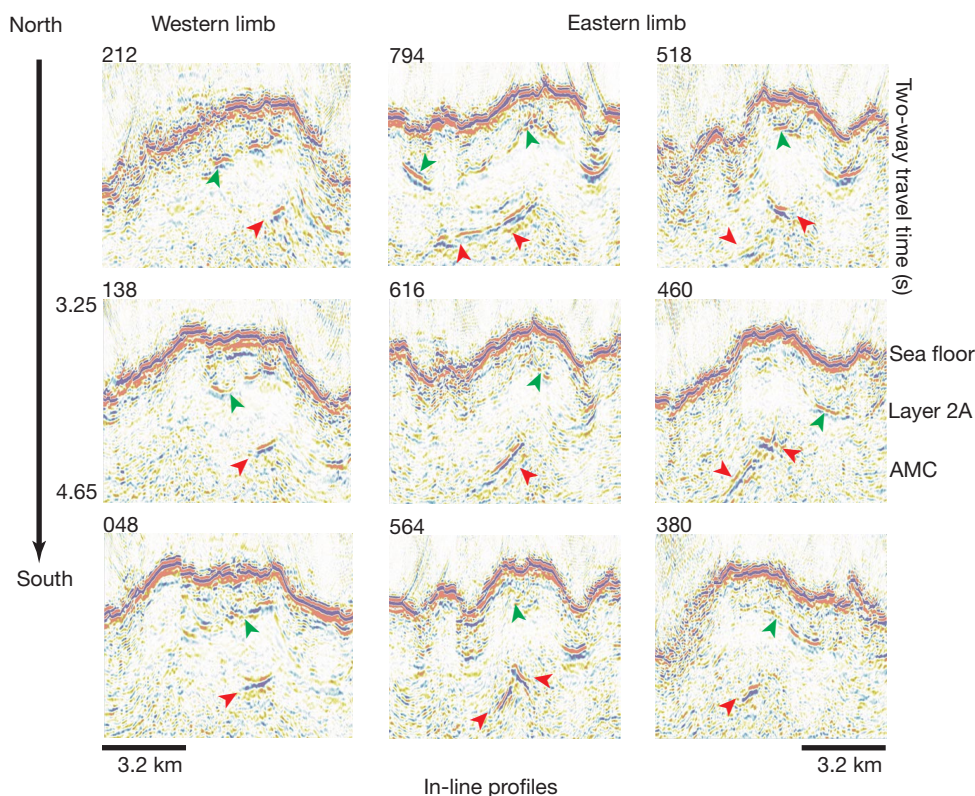


Figure 2 In-line seismic reflectivity slices through the western and eastern limbs of the 9° 03' N overlapping spreading centre. Axial magma chambers and seismic layer 2A reflectors are highlighted with red and green arrows, respectively. The three-dimensional melt sill structure beneath the western limb is relatively simple, showing a gradual decrease in melt sill width from 1,100 m (slice 048) beneath an inflated axial ridge, to less than 375 m (slice 212) some 4 km northwards, just before the melt sill event disappears (see Fig. A1 in Supplementary Information). In contrast to the western limb, the melt sill structure beneath the eastern limb is very complex. To the north, the eastern limb is dominated by a westward-displaced melt sill (slice 794), which underlies the overlap basin. This feature shows little change for 4 km along-strike, then its width quickly tapers down to <1,500 m (slice 616). Southwards, a second melt sill event is seen dipping toward the east, producing a 'tent-shaped' magma chamber morphology (slice 564).

Next, a ridge-centred sill is observed (slice 518), with an intermittent, secondary melt sill extending from the central overlap basin, upwards toward the primary sill (slice 460). These secondary events demonstrate that the ridge-centred melt sill beneath the eastern limb is fed, in part, from magma upwelling beneath the overlap basin. The ridge-centred melt sill continues to deepen (slice 380) towards the propagating ridge tip; despite its narrow width (< 500 m), this event is nearly continuous along the axis, until it finally dies out near the propagating tip. Time-slice seismic reflectivity maps through the western and eastern limbs are available; see Fig. A1 in Supplementary Information. Slice-by-slice cross-dip views of the melt sill beneath both limbs are available as Supplementary Information movies "In-line Western Limb" and "In-line Eastern Limb". AMC, axial magma chambers.

process and coupling of magmatic, tectonic and petrologic segmentation of the ridge crest. The three-dimensional nature of melt distribution across this offset (Fig. 4a) appears to be inconsistent with large-scale, crustal redistribution of melt away from centres of upwelling some 50–100 km distant^{2,3}; there are two reasons why this is the case. First, the melt sill along the propagating limb has a tortuous morphology with contributions from magma sources beneath the overlap basin. This geometry is hard to explain by simple along-axis flow away from a distant centre of upwelling, as such flow would require moving buoyant material to deeper levels within the crust, with a significant off-axis component of deflection toward the overlap basin. Second, it is interesting to note that the deepening of the axial magma chamber away from the inferred centre of upwelling is approximately 250 m over the first 90 km (refs 15, 25), but then plunges twice that amount over the last 6 km. The increased gradient over the last 6 km is most probably related to variations in the rheology and thermal structure of the overlying

upper crust at the propagating southern tip of the eastern limb, although this depth increase may be related to the injection of melt into the propagator²⁶, and the associated hydraulic head loss within the reservoir. The latter scheme is more plausible if a sustaining low-velocity zone is either muted or non-existent beneath the southern tip. Interestingly, the magma reservoir makes several small (200 m)

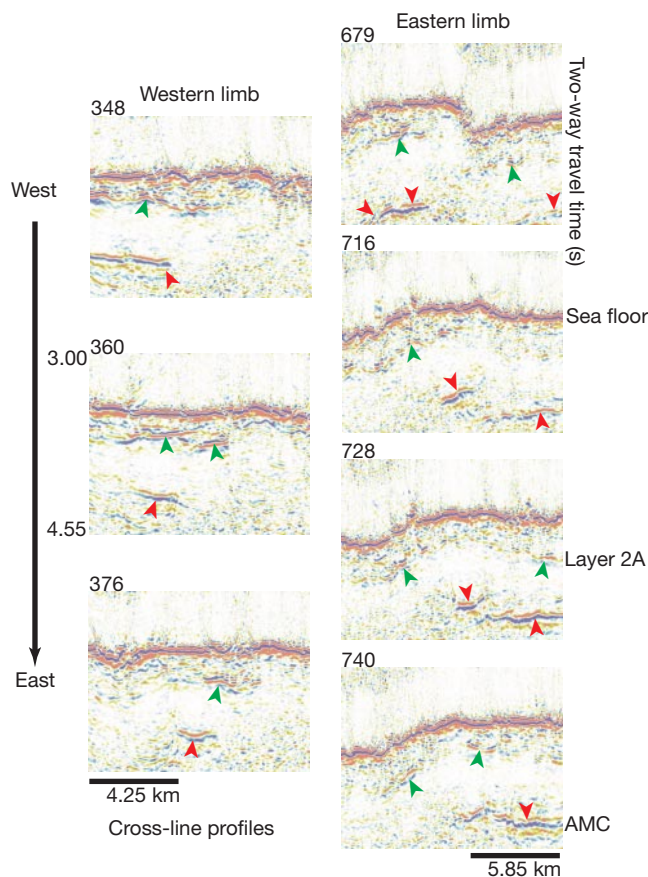


Figure 3 Cross-line seismic reflectivity slices through the western and eastern limbs of the 9° 03' N overlapping spreading centre. Axial magma chambers and seismic layer 2A reflectors are highlighted with red and green arrows, respectively. Three cross-line slices illustrate the simpler axial geometry of the melt lens beneath the western limb, although changes in layer 2A thickness might control (assuming constant-thickness layer 2B) the exact location of the melt sill (slice 360) beneath the ridge crest. Four in-line slices through the eastern limb (and overlap basin) show an abrupt decrease in magma chamber depth (slices 716 and 728) moving southwards along the propagating ridge tip, where melt sill morphology changes from a westward-displaced event into a ridge-centred system. Further along the propagating tip (southwards), the melt horizon shows a significant increase in depth relative to either sea-floor or layer 2A events (roughly 500 m depth over 6 km of ridge length), indicative of rapid changes in thermal, rheology and permeability structure beneath the propagating tip of this limb. Supplementary Information movies "Cross-line Western Limb" and "Cross-line Eastern Limb" allow a slice-by-slice along-strike view of the melt sill beneath both limbs.

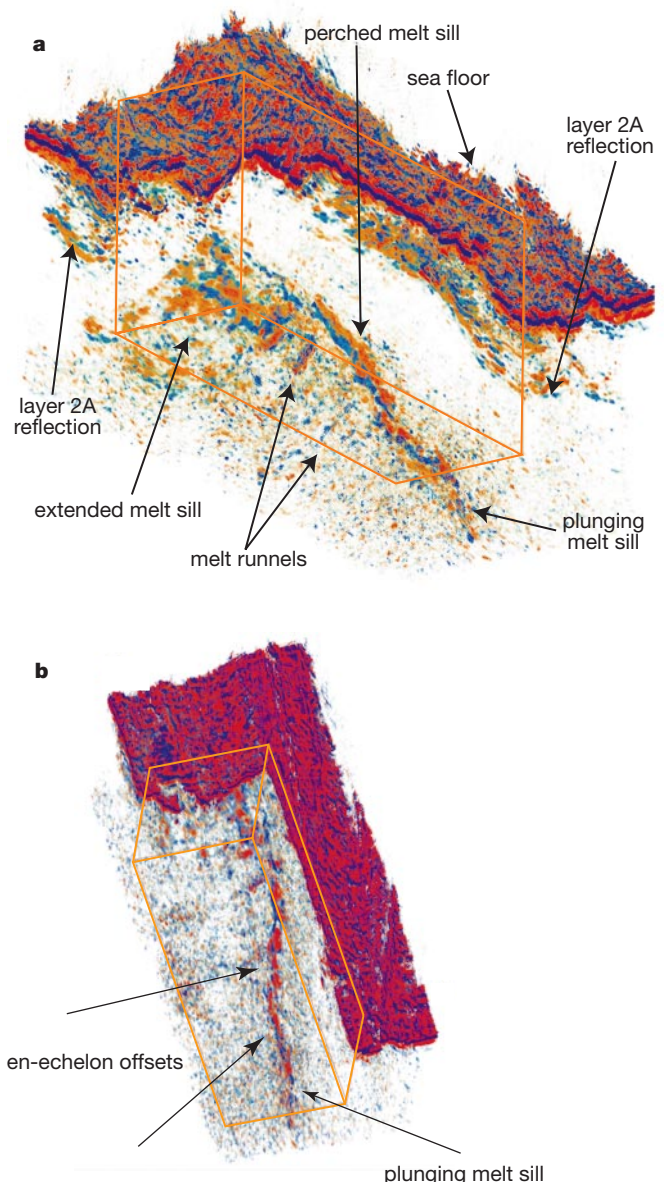


Figure 4 Three-dimensional visualization of magma chamber structure. **a**, Three-dimensional rendering of seismic reflectivity beneath the eastern limb of the 9° 03' N overlapping spreading centre, East Pacific Rise. A corner-cut (orange frame) has been made to allow viewing of the magma chamber reservoir (orange and blue-green hues), through the sea-floor and layer 2A reflections (red and blue hues). This oblique perspective of the magma chamber structure highlights the abundance of melt beneath both the northern overlap basin and the propagating limb of this large ridge-axis discontinuity. The complex melt sill morphology underlying the eastern limb is inconsistent with large-scale, along-axis crustal flow or migration of magma from a distant centre of upwelling. **b**, A close-up view of fine-scale magma chamber structure beneath the propagating eastern limb. Note the en-echelon behaviour of the plunging melt sill, which suggests melt migration within mid-crustal fractures or cracks. Relevant movies are available as Supplementary Information; 'ARAD_Infill' and 'ARAD-Spin' for **a**, 'ARAD_Tilt' for **b**.

right-lateral steps (Fig. 4b) within the zone of deepening, which is suggestive of melt injection through en-echelon, mid-crustal cracks or fractures. This mode of rift-tip injection might also explain the asymmetry of melt between the two limbs.

Lavas erupted near the propagating tip of this offset²⁷ were found to have high concentrations of titanium and iron oxides, which led investigators to suggest that these Fe-Ti basalts were derived from isolated magma bodies within the rift-tip¹. The largest magnetic anomaly in this region²⁷, delimiting the inferred zone of Fe-Ti basalts, coincides with the downward-plunging section of the magma chamber. The current melt sill morphology rules out an isolated melt body, although the en-echelon structure of the plunging melt sill would probably reduce any along-axis mixing of magma. Thus, the resultant geochemistry may better reflect the thermal environment within the propagating limb, or possibly a short-wavelength, along-axis fractionation trend²⁸.

Estimates of crustal²⁹ and extrusive¹⁷ thickness surrounding the overlap region are significantly greater than observed near 9° 50' N, which is the inferred focus of magma supply for this segment^{2,3}. If the current mode of melt emplacement along this ridge segment has remained more or less constant on a timescale of 10–100 kyr, or remained stationary with respect to the propagating offset, then it is unlikely that any large-scale crustal redistribution of melt is responsible for thickened crust throughout this region, although sub-crustal migration and focusing of melt cannot be ruled out. An alternative, simpler explanation requires only small-scale (a few tens of kilometres) redistribution of melt along the ridge crest^{10–12}, which could be further modified through focusing along permeable pathways within the middle to upper crust. □

Methods

Navigation and binning

The hydrophone array used during the ARAD experiment was a 3,100-m-long, 124-channel Digicon DMS-2000 streamer. Shipboard navigation was achieved by an Inmarsat-based differential GPS system with an accuracy of a few metres. Streamer positioning was reconstructed using bearing measurements (with an accuracy of a few tenths of a degree) from 12 Digicon compass birds placed at 250 m intervals along the streamer. Streamer positioning was checked against GPS fixes from the streamer tailbuoy, which confirmed our solution with along- and cross-track errors no greater than 15 and 30 m, respectively. Individual shot records from a 10-gun (50-litre) source array were spaced 38 m apart, and were binned according to common-midpoint location between source and receiver. To achieve even offset coverage, a flexible binning strategy was adopted whereby individual bins (25 m × 100 m) were overlapped by 50% in the cross-line direction. Bin extension lessens the effect of uneven coverage at a given source–receiver range, which is rooted in the variability of streamer shape throughout the experiment. Eight additional profiles were collected to fill in the most serious gaps in data coverage.

Stacking and migration

Three-dimensional images of seismic reflectivity were obtained through the analysis of some 120,000 shot records. The > 13 million binned traces were spike-edited, filtered, corrected for spherical divergence and normal moveout, and stacked. The resultant 20 km × 20 km stacked volume was first in-line migrated at the water velocity to collapse steeply dipping diffractions originating at the sea floor, followed by cross-line interpolation onto 25 m × 25 m cells. The interpolated data volume was then cross-line migrated at the water velocity to collapse any remaining sea-floor-diffracted energy. Lastly, the entire data volume was finite-difference time-migrated (two-pass) using residual velocities³⁰ to produce a geometrically correct version of the sea floor and underlying features such as melt sills, extrusive layer-dike contacts and Moho. The interpretations presented here are based on the migrated three-dimensional time-image, with conversions to depth, where given, based on a composite velocity function¹⁷ that accounts for lateral variations in layer 2A thickness. The mapping of events into depth based on this scheme assumes that deepening of the 2A interface in travel time can be, to first order, accounted for through an increase in thickness and velocity of the surficial layer, along with a systematic relaxation of the velocity gradient at the base of layer 2A¹⁷.

Received 29 November 1999; accepted 9 June 2000.

- Sinton, J. M., Wilson, D. S., Christie, D. M., Hey, R. N. & Delaney, J. R. Petrologic consequences of rift propagation on oceanic spreading ridges. *Earth Planet. Sci. Lett.* **62**, 193–207 (1983).
- Macdonald, K. C. *et al.* A new view of the mid-ocean ridge from the behaviour of the ridge-axis discontinuities. *Nature* **335**, 217–225 (1988).
- Wang, X., Cochran, J. R. & Barth, G. Gravity anomalies, crustal thickness and the pattern of mantle flow at the fast spreading East Pacific Rise, 9°N–10°N: Evidence for three-dimensional upwelling. *J. Geophys. Res.* **101**, 17927–17940 (1996).

- Lonsdale, P. Segmentation of the Pacific-Nazca spreading center, 1°N–20°S. *J. Geophys. Res.* **94**, 12197–12225 (1989).
- Macdonald, K. C. & Fox, P. J. Overlapping spreading centres: new accretion geometry on the East Pacific Rise. *Nature* **302**, 55–58 (1983).
- Lonsdale, P. Overlapping rift zones at the 5.5°S offset of the East Pacific Rise. *J. Geophys. Res.* **88**, 9393–9406 (1983).
- Langmuir, C. H., Bender, J. F. & Batiza, R. Petrologic and tectonic segmentation of the East Pacific Rise, 5°30'–14°30'N. *Nature* **322**, 422–429 (1986).
- Scheirer, D. S. & Macdonald, K. C. Variation in cross-sectional area of the axial ridge along the East Pacific Rise: Evidence for the magmatic budget of a fast spreading center. *J. Geophys. Res.* **98**, 7871–7886 (1993).
- Magde, L. S., Detrick, R. S. & TERA Group. Crustal and upper mantle contribution to the axial gravity anomaly at the southern East Pacific Rise. *J. Geophys. Res.* **100**, 3747–3766 (1995).
- Toomey, D. R., Purdy, G. M., Solomon, S. C. & Wilcock, W. S. D. The three-dimensional seismic velocity structure of the East Pacific Rise near latitude 9°30'N. *Nature* **347**, 639–645 (1990).
- Singh, S. C., Kent, G. M., Collier, J. S., Harding, A. J. & Orcutt, J. A. Melt to mush variations in crustal magma properties along the ridge crest at the southern East Pacific Rise. *Nature* **394**, 874–878 (1998).
- Dunn, R. A. & Toomey, D. R. Seismological evidence for three-dimensional melt migration beneath the East Pacific Rise. *Nature* **388**, 259–262 (1997).
- Mutter, J. C. *et al.* Magma distribution across ridge axis discontinuities on the East Pacific Rise (8°40' to 9°50' North) from multichannel seismic images. *Nature* **336**, 156–158 (1988).
- Collier, J. S. & Sinha, M. C. Seismic images of a magma chamber beneath the Lau Basin back-arc spreading center. *Nature* **346**, 646–648 (1990).
- Kent, G. M., Harding, A. J. & Orcutt, J. A. Distribution of magma beneath the East Pacific Rise near the 9°03'N overlapping spreading center from forward modeling of common depth point data. *J. Geophys. Res.* **98**, 13971–13996 (1993).
- Carbotte, S. & Macdonald, K. C. East Pacific Rise 8°–10°30'N: Evolution of ridge segments and discontinuities from SeaMARC II and three-dimensional magnetic studies. *J. Geophys. Res.* **97**, 6959–6982 (1992).
- Harding, A. J., Kent, G. M. & Orcutt, J. A. A multichannel seismic investigation of the upper crustal structure at 9°N on the East Pacific Rise: Implications for crustal accretion. *J. Geophys. Res.* **98**, 13925–13944 (1993).
- Carbotte, S., Mutter, J. C. & Wu, L. Contribution of volcanism and tectonism to axial and flank morphology of the southern East Pacific Rise, 17°10'–17°40', from a study of layer 2A geometry. *J. Geophys. Res.* **102**, 10165–10184 (1997).
- Christeson, G. L., Kent, G. M., Purdy, G. M. & Detrick, R. S. Extrusive thickness variability at the East Pacific Rise, 9°–10°N: Constraints from seismic techniques. *J. Geophys. Res.* **101**, 2859–2873 (1996).
- Hooff, E. E. E., Schouten, H. & Detrick, R. S. Constraining crustal emplacement processes from the variation of seismic layer 2A thickness at the East Pacific Rise. *Earth Planet. Sci. Lett.* **142**, 289–309 (1996).
- Kent, G. M., Harding, A. J. & Orcutt, J. A. Evidence for a smaller magma chamber beneath the East Pacific Rise at 9°30'N. *Nature* **344**, 650–653 (1990).
- Bazin, S. *Three-dimensional Crustal Structure of East Pacific Rise Discontinuities from Tomographic Inversions*. Thesis, Univ. California, San Diego (2000).
- Singh, S. C., Collier, J. S., Harding, A. J., Kent, G. M. & Orcutt, J. A. Seismic evidence for a hydrothermal layer above the solid roof of the axial magma chamber at the southern East Pacific Rise. *Geology* **27**, 219–222 (1999).
- Phipps Morgan, J. & Chen, Y. J. The genesis of oceanic crust: Magma injection, hydrothermal circulation, and crustal flow. *J. Geophys. Res.* **98**, 6283–6297 (1993).
- Detrick, R. S. *et al.* Multichannel imaging of the axial magma chamber along the East Pacific Rise between 9°N and 13°N. *Nature* **326**, 35–41 (1987).
- Phipps Morgan, J. & Parmentier, E. M. Causes and rate-limiting mechanisms of ridge propagation: a fracture mechanics model. *J. Geophys. Res.* **90**, 8603–8612 (1985).
- Sempère, J.-C., Macdonald, K. C. & Miller, S. P. Overlapping spreading centres: 3-D inversion of the magnetic field at 9°03'N on the East Pacific Rise. *Geophys. J. R. Astron. Soc.* **79**, 799–811 (1984).
- Batiza, R. & Mutter, J. C. Variability in oceanic crustal thickness and structure: Multichannel seismic reflection results from the northern East Pacific Rise. *J. Geophys. Res.* **101**, 17951–17975 (1996).
- Barth, G. A. & Mutter, J. C. Variability in oceanic crustal thickness and structure: Multichannel seismic reflection results from the northern East Pacific Rise. *J. Geophys. Res.* **101**, 17951–17975 (1996).
- Yilmaz, O. *Seismic Data Processing* (Society of Exploration Geophysicists, Tulsa, 1987).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or in CD-ROM format from the author.

Acknowledgements

We thank the officers and crew of the RV *Maurice Ewing* for around-the-clock help during the 47-day-long cruise; H. Avendonk, D. Boschi, A. Cherrett, A. Greer, P. Henkart, C. Hollinshead, S. Hussenoeder, T. Owen and P. Zimmer for assistance with instrumentation and watch-standing duties during the experiment; and R. Detrick and D. Toomey for comments that significantly improved this Letter. Data-processing software was provided by Paradigm Geophysical, Houston, Texas. Focus 3-D was used for the three-dimensional binning, stacking and migration of the multichannel seismic data. Data presentation of the three-dimensional volume was facilitated through visualization modules within Focus 3-D and VoxelGeo. The ARAD seismic experiment is an international collaborative project between investigators at the Scripps Institution of Oceanography and the University of Cambridge, with financial support from the RIDGE program/National Science Foundation (USA), British Institutions Reflection Profiling Syndicate (BIRPS), and the Natural Environment Research Council (UK). Additional matching funds for computer visualization and data storage were provided by the Scripps Institution of Oceanography and the Cecil H. and Ida M. Green Foundation for Earth Sciences.

Correspondence and requests for materials should be addressed to G.M.K. (e-mail: gkent@ucsd.edu).

Fine structure of bone in dinosaurs, birds and mammals

John M. Rensberger* & Mahito Watabe†

* Burke Museum of Natural History and Culture & Department of Geological Sciences, University of Washington, Seattle, Washington 98195, USA

† Hayashibara Museum of Natural Sciences, Shimoishii 1-2-3 Okayama 700-0907, Japan

After observation of detailed structural evidence for the origin of birds from dinosaurs¹, and in light of evidence that dinosaur bone tissue resembles the histology in mammals², the histology of bone has become one of the focal points in discussions of the physiology of dinosaurs and Mesozoic birds^{3–10}. Most of this microstructural information has focused on features related to the vascular organization and the amount of remodelled bone around vascular canals. However, the finer structures have received less attention, although differences in such structures have been observed among modern vertebrates^{10,11}. Here we present evidence that canaliculi—the submicrometre-sized channels that interconnect bone cells and vascular canals—and the collagen fibre bundles in bone are differently organized among certain dinosaur lineages. Ornithomimid dinosaurs¹² are more like birds than mammals in these features. In canalicular structure, and to some extent in fibre bundle arrangement, ornithischian dinosaurs are more like mammals. These differences in both canalicular and lamellar structure are probably linked to differences in the process and rate¹³ of bone formation.

The canalicular organization in mammals is distinct from that in birds and ornithomimid dinosaurs (Fig. 1). In mammalian bone, other than in the fast-growing tissue of very young individuals or of fracture calluses, the canaliculi are typically aligned in parallel with the direction in which bone deposition proceeds, that is, in a radial direction toward either the outer, periosteal surfaces of the bone or the inner, endosteal surfaces of medullary spaces, or the lumina of

vascular canals^{10,11}. When canaliculi divide, the branches tend to run almost parallel to the parent canaliculus, in a radial orientation¹⁴. We have found this to be the main canalicular orientation in adult cortical bone of the ten species (six orders) of mammals that we have studied; quantitative data show the limited extent to which the canaliculi diverge from the radial direction (Figs 2a and 3).

More irregular canalicular structure occurs in birds^{10–11} but the structure and its distribution have been poorly documented. Our

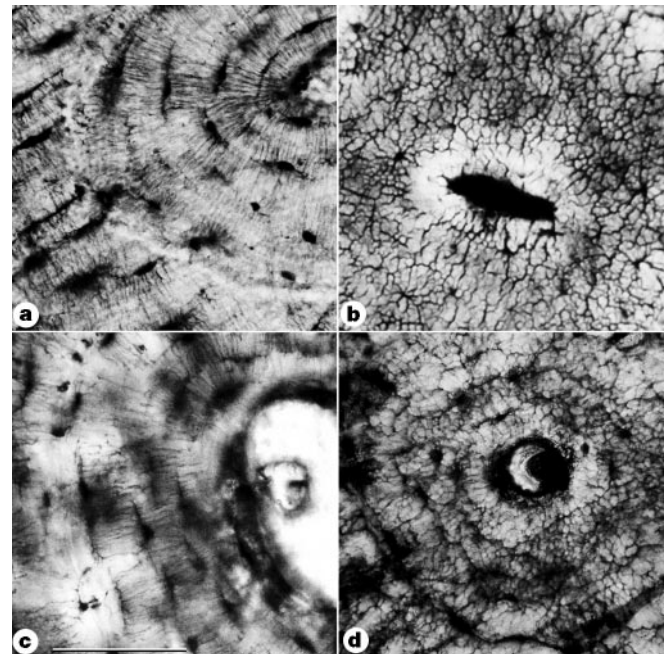


Figure 2 Canalicular organization in transverse thin sections of various bone types. **a**, Barbary sheep humerus, secondary osteon in periosteal bone; **b**, bald eagle tarsometatarsus, secondary osteon adjacent to endosteal region; **c**, hadrosaur tendon, detail of secondary osteon (UWBM (Burke Museum) 76357a); **d**, ornithomimid metacarpal, secondary osteon (UWBM 76332). Scale bar, 100 μ m.

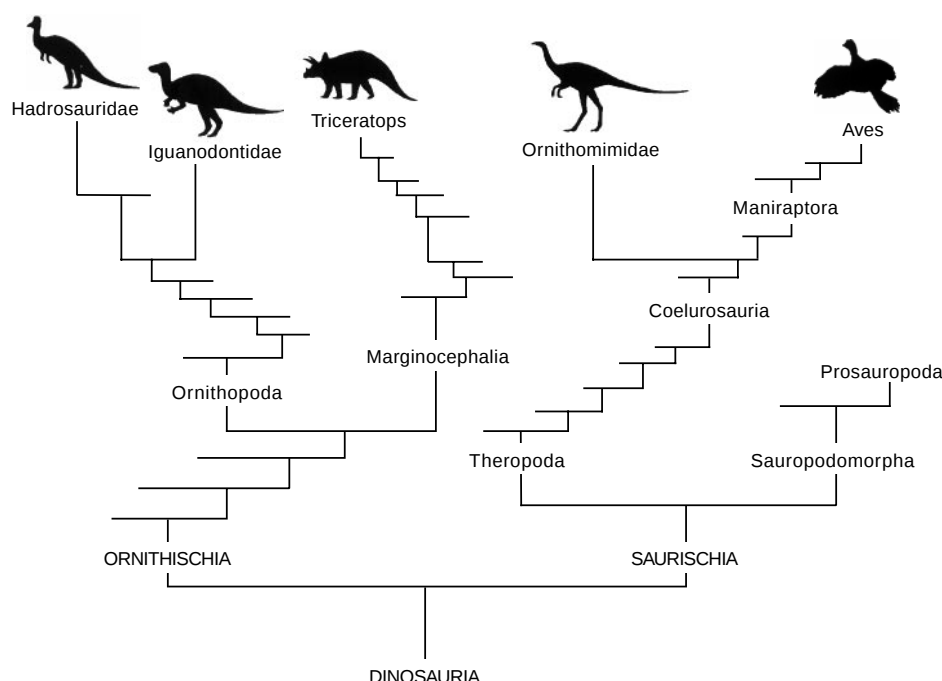


Figure 1 Relationships of certain dinosaur taxa. Modified from ref. 12.

examination of tissue samples from the major limb bones in 17 species (11 orders) of birds of diverse adaptations show that the canaliculi in adults are consistently organized throughout the bone shafts as extensively branching channels that diverge at large angles from the parent canaliculus (Fig. 2b), making the directional pattern essentially random (Fig. 3).

In spite of their small diameters, canaliculi are often preserved in fossil bones, and photographic images have been published^{7,15}, although only extending for short distances from bone cell lacunae. Two dinosaur specimens from the late Cretaceous Hell Creek Formation in North America, both medium-sized ornithomimids (Fig. 1), have canaliculi organized throughout the interlacunar space in a randomly branching network (Fig. 2d) that is indistinguishable in geometry (Fig. 3) and dimensions from the pattern in birds. Although more faintly preserved, the canalicular organization is also recognizable in several theropod dinosaurs from the later Cretaceous Nemegt Formation in Mongolia. In two specimens of the ornithomimid *Gallimimus* and in a specimen of another small theropod, the structure is essentially the same random network that characterizes the two North American ornithomimids. This structure probably also characterizes coelurosaurs other than ornithomimids, considering the evidence for derivation of birds from a separate lineage of coelurosaurs (Fig. 1)^{16,17}.

In contrast to the canalicular pattern in extant birds and the ornithomimids, the canaliculi in the ornithischian dinosaurs that we have examined are quite different. Canaliculi in limb bone fragments of a hadrosaur from the Nemegt Formation have the radial, parallel alignment that is characteristic of mammals. Canaliculi in sections of the frill of *Triceratops* and from ossified hadrosaur tendon (Fig. 2c) from the Hell Creek Formation have the same structure.

Another histologic feature, the spatial organization of the collagen fibre bundles^{10,18}, is associated with the canalicular differences in bones. Bundles of apatite crystallites, which are intimately associated with collagen in bone, have been observed with electron

microscopy on free surfaces of dinosaur bone—such as the surfaces of vascular canals and lacunae¹⁹—and striated collagen fibrils have been seen in decalcified dinosaur bone²⁰ with transmission electron microscopy. These observations do not reveal the general fibrillar organization through the thickness of the tissue. However, thin sections of dinosaur bone under polarized light microscopy (PLM) have often exhibited a pattern of alternating light and dark birefringent bands, resulting from differential rates of light transmission dependent on crystal orientation, that is characteristic of the lamellar bone of modern vertebrates^{3,6}.

Under PLM, individual lamellae in both primary and secondary (replacement) osteons (cylinders of bone deposited around vascular canals) in mammals are usually well defined and continue over considerable distances (Fig. 4a); statistical analysis shows that they have relatively uniform thicknesses (Fig. 5). This regularity is seen in endosteal lamellar bone, in primary osteons where developed around circumferentially extended vascular canals and in non-osteonal, circumferential primary lamellae in the outer part of the cortex.

In contrast, individual endosteal lamellae in birds are irregularly defined^{10,11}, tend to narrow, disappear or divide over short distances (Fig. 4b), and vary in thickness (Fig. 5). Osteonal lamellae are usually lacking in birds, but when present they are irregular in thickness and disappear within short intervals. Similarly, in the North American and Asian ornithomimids that we examined, lamellae in general are irregular in thickness, taper and disappear, as in birds and unlike the continuous lamellae of relatively constant thicknesses in mammals. This is true of tracts of endosteal lamellae bordering the medullary cavities (Fig. 4d), as well as osteonal lamellae where they occur.

However, in the ornithischians (hadrosaurids, including *Sauroplophus*; *Protoceratops* and *Triceratops*), many of the secondary osteons have several continuous lamellae with relatively constant thicknesses (Fig. 4c), as in mammals. In contrast, in primary osteonal lamellae of the cortex, the organization of the fibrillar structure in the ornithischians may be more irregular, resembling that in the ornithomimids and birds. These observations for the ornithischians are consistent with those made by Currey³ for a sauropodomorph dinosaur. He illustrated and described tissue in the primary bone of a prosauropod dinosaur as having (under

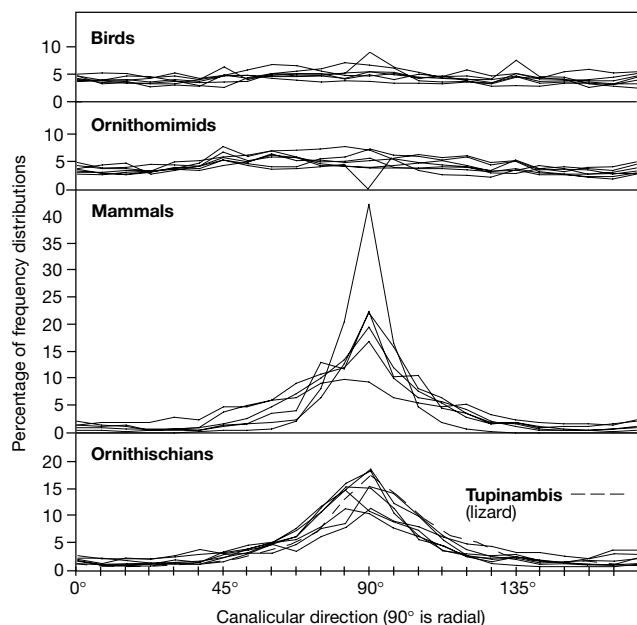


Figure 3 Percentage frequency distributions of canalicular directions in samples of birds, ornithomimids, mammals and ornithischians. Birds: we used a bald eagle phalanx, an emu femur, a Siberian swan tibiotarsus, and a Canada goose femur. Ornithomimids: we used the UWBM 31778 phalanx and the UWBM 76332 metacarpal. Mammals: we used a pronghorn metatarsal and a puma femur. Ornithischians: we used the hadrosaur UWBM 76357 tendon and the *Triceratops* UWBM 77379 squamosal. A lizard (*Tupinambis* sp. humerus) is included with the ornithischians for comparison. An entirely random frequency distribution would be represented by a horizontal line at 4.17%.

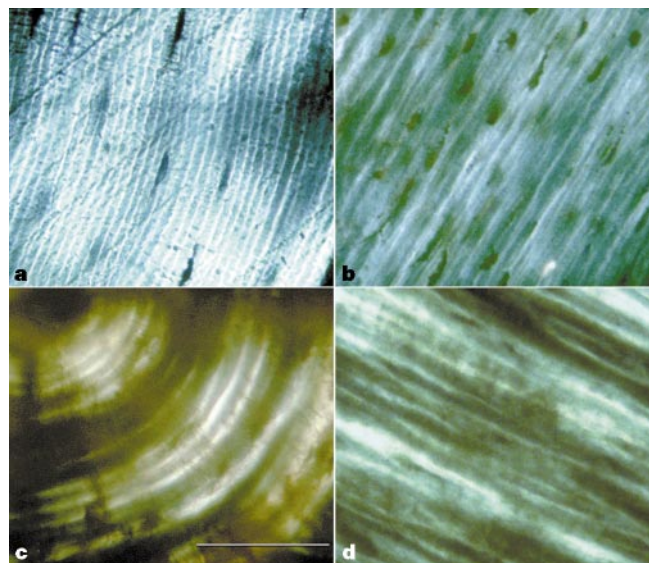


Figure 4 Lamellar organization as seen with crossed polars in transverse thin sections of various bones. **a**, Straw-coloured fruit bat ulna, periosteal; **b**, Canada goose femur, endosteal; **c**, Hadrosaur tendon UWBM 76357b, secondary osteon; **d**, *Gallimimus* metatarsal, Hayashibara Museum (HMNS) 940813, endosteal. Scale bar, 100 μ m.

PLM) "bright and dark regions, but no true lamellae in the primary osteons". He observed that this was probably not due to poor preservation because in adjacent tissue he found secondary osteons containing distinct, well developed lamellae.

There are indications in birds and ornithomimids of positional correlation of the degree of lamellar organization with the degree of canalicular organization. Primary and secondary osteons in both groups consist mainly of parallel-fibred bone with full-bodied osteocyte lacunae. However, a few flattened osteocyte lacunae, with long axes aligned circumferentially with respect to the vascular canal, sometimes occur locally near the outer perimeter of the osteon or adjacent to the vascular canal. A few radial canaliculi are often present in these regions (Fig. 2b and d), and in some cases a pair or more of alternately bright and dark lamellar bands occur there. In the main body of the osteons, however, the lacunae tend to be cylindrical and lack the compressed, lenticular shape, and under PLM the tissue shows little or no indication of lamellae. Also, in the endosteal region in birds, where lamellae are often present, canaliculi, although irregular, have a directed, radial orientation toward the marrow cavity. The association in both these regions of lamellar structure and canaliculi showing some degree of parallel orientation indicates a close developmental relationship between the organization of the canalicular structure and the fibrillar structure.

Amprino²¹ predicted a correspondence between the regularity of the fibrillar organization and growth rate. Lamellar bone, which has the highest degree of fibrillar organization in mammals^{18,22–24}, has the slowest rate of growth, whereas woven-fibred bone, which has the poorest degree of fibre organization, has the highest rate of formation; parallel-fibred bone has an intermediate rate^{13,25–27}. This suggests that tissues with poorly organized lamellar structure and randomly directed canaliculi were deposited faster than tissues with well defined, regular lamellae and orientated, parallel canaliculi. Lamellar bone is scarce in birds and coelurosaurs, and where it does occur it is poorly organized. In these taxa, much of the skeleton is restricted to more rapidly forming parallel-fibred bone; when lamellar bone is present, its more irregular organization suggests that it was deposited at higher rates than lamellar bone in mammals.

A geometric strategy to increase the rate of skeletal growth in taxa that must reach large size rapidly is the formation of multiple surfaces to receive subsequent bone. Bone formed in this way, sometimes called plexiform tissue, is initially deposited as rapidly

growing woven-fibred or parallel-fibred bone in trabeculae that enclose relatively large vascular spaces; these become filled in later and more slowly with lamellar bone of the primary osteons²⁸. Because lamellar bone must be deposited on pre-existing surfaces, the greater surface areas provided by the trabecular extensions allow a greater quantity of bone to be deposited simultaneously. However, the completion of bone formation in this tissue awaits the filling of the vascular spaces by lamellar bone. In dinosaurs and birds that have plexiform tissues, the greater irregularity of the lamellae suggests that this process is more rapid than in the plexiform tissues of mammals.

The vascular²⁵ as well as fibrillar²¹ structure of bone tissue can change between early and mature stages of growth in relation to reduction in appositional rates. How the structural conditions for ornithomimid, ornithopod and ceratopsian dinosaurs varies across ontogenetic series from young to old individuals will remain uncertain until samples from individuals in a single taxon that are obviously quite different in age are compared. However, the various ornithomimids that we have examined are remarkably consistent in their canalicular organization, even when comparing primary bone from the outer cortex with bone of secondary osteons, which are tissues of different ontogenetic age. The consistency of the canalicular and fibrillar structure in the substantial samples of modern adult birds adds to the predictive value of the smaller samples of dinosaur bone tissue.

The predominance of randomly directed canaliculi and the irregular organization of collagen bundles in birds and ornithomimids thus predicts that a more extensive distribution of these features will be found among other coelurosaurs. Highly organized lamellar bone and radially directed canaliculi may be the primitive condition for tetrapods. We confirm earlier observations of directed, parallel canaliculi in modern amphibians^{11,29}. This, together with the preponderance of well organized lamellae and canalicular structure in mammals and probably squamate reptiles (Figs 3 and 5), suggests that the more irregular structural organization characteristic of coelurosaurs and birds, and of tissue-specific regions in other dinosaurs, is a derived condition. The selective value of this innovation appears to be high rates of growth of the bones, a metabolically expensive strategy that would be associated with elevated rates of nutrient acquisition and structural innovations to conserve energy, such as the filamentous integumentary structures recently found in several non-avian coelurosaurs³⁰. □

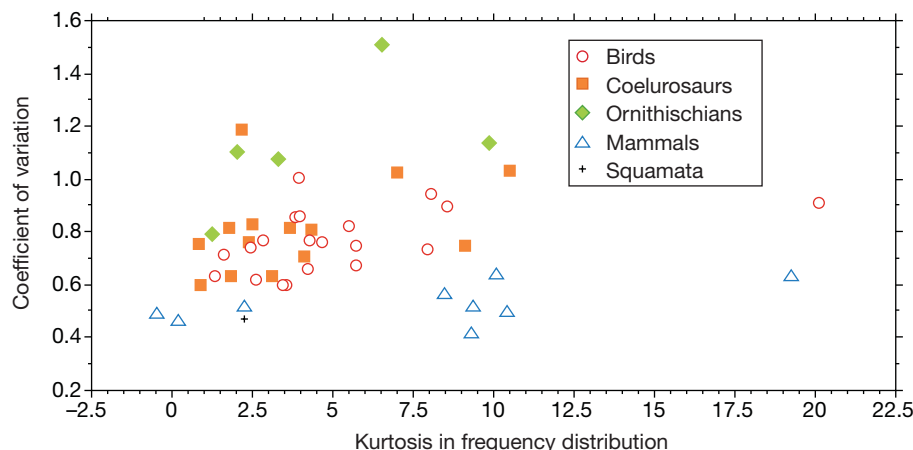


Figure 5 Degree of variability of lamellar thicknesses in samples of mammals, birds, coelurosaurs and lizard. Mammals: we used puma, Barbary sheep, Townsend mole, straw-coloured fruit bat. Birds: we used Canada goose, Pacific loon, red-tailed hawk, bald eagle, raven. Coelurosaurs: we used *Gallimimus* HMNS 940813, small Mongolian theropod HMNS 940810; North American ornithomimid UWBM 31778; ornithischian (*Iguanodon* HMNS 960718). Squamata: lizard, *Tupinambis* sp. Each coefficient of

variation is based on 2,000 to 15,000 thickness measurements of dark lamellar bands bounded by light bands in a region varying from 7,326 to 17,056 μm^2 . Coefficients of variation are lower in the mammals than in the birds and dinosaurs ($P = 0.00002$; Fisher's Exact test). Kurtosis, a measure of peakedness of the frequency distribution, is greater in the mammals ($P = 0.0034$).

Methods

Thin sections

We cut sections 0.7–1.1 mm thick from fragments of the fossil bones with a Leitz Saw Microtome and ground them to 70 µm or less. Sections of Recent bones were cut 1 mm thick with a scroll saw and ground to thicknesses ranging from 20 to 100 µm.

We examined, with transmitted light microscopy and PLM, more than 500 optical thin sections of Recent taxa, taken from several recorded positions along the diaphyses of limb bones in each species. In birds, we sampled the humerus, ulna, femur, tibiotarsus, tarsometatarsus and phalanges in adults of the following orders and species: Struthioniformes: *Dromaius novaehollandiae* (emu); Galliformes: *Meleagris gallopavo* (turkey) and *Phasianus colchicus* (ring-necked pheasant); Anseriformes: *Cygnus columbianus* (tundra swan), *Branta canadensis* (Canada goose) and *Anas strepera* (gadwall); Gaviiformes: *Gavia pacifica* (Pacific loon); Podicipediformes: *Podiceps griseigena* (red-necked grebe); Pelecaniformes: *Pelecanus erythrorhynchos* (white pelican); Ciconiiformes: *Ardea herodias* (great blue heron); Charadriiformes: *Larus delawarensis* (ring-billed gull); Falconiformes: *Haliaeetus leucocephalus* (bald eagle) and *Buteo jamaicensis* (red-tailed hawk); Strigiformes: *Tyto alba* (barn owl) and *Bubo virginianus* (great horned owl); Passeriformes: *Corvus corax* (common raven) and *Carduelis pinus* (pine siskin). We also sampled bones from a nestling *Tyto alba* and a fledgling *Accipiter striatus* (sharp-shinned hawk). We examined approximately 220 thin sections from the major limb bones in adults of the following orders and species of mammals: Marsupialia: *Didelphis marsupialis* (opossum) and *Macropus rufus* (red kangaroo); Insectivora: *Scapanus townsendii* (Townsend mole); Chiroptera: *Eidolon helvum* (straw-coloured fruit bat); Rodentia: *Sciurus griseus* (grey squirrel) and *Capromys pilorides* (hutia); Carnivora: *Procyon lotor* (raccoon) and *Puma concolor* (puma); Artiodactyla: *Antilocapra americana* (pronghorn) and *Ammotragus lervia* (Barbary sheep).

Measurements of canalicular directions

The measurements of canalicular directions are based on examples with typical as well as more extreme variations in structure where the canaliculi show good contrast over an extended area. We scanned film images made at a microscope magnification of 400× and locally enhanced the image contrast of the canaliculi, then converted the images to black and white bitmaps. The directions of canalicular boundaries in the bitmaps were calculated using software that measures a direction from each pixel along the boundary to a position six pixels ahead on the same boundary, repeats this for all canalicular edges and prepares a frequency distribution of the directions throughout the image.

Measurements of lamellar thicknesses

Using PLM film images, we measured the degree of uniformity of lamellar thickness in areas with best lamellar contrast. Although we measured lamellae in endosteal, osteonal and periosteal bone tissue, the data shown in Fig. 5 exclude osteonal lamellae because relatively straight lamellae could be measured more exhaustively and accurately by software. We high-pass filtered each digital sample identically to eliminate low-frequency contrast across the image, then increased the contrast of the resulting greyscale image and converted it to a bitmap. The software calculated the distance from one lamellar boundary to the next boundary in the direction perpendicular to the general direction of the lamellar planes, and repeated the process for each pixel along each lamellar boundary.

Received 21 October 1999; accepted 18 May 2000.

- Ostrom, J. H. The ancestry of birds. *Nature* **242**, 136 (1973).
- Enlow, D. H. & Brown, S. O. A comparative histological study of fossil and recent bone tissue. Part II. *Tex. J. Sci.* **9**, 186–214 (1957).
- Currey, J. D. The histology of the bone of a prosauropod dinosaur. *Palaeontology* **5**, 238–246 (1962).
- Bakker, R. T. Anatomical and ecological evidence of endothermy in dinosaurs. *Nature* **281**, 81–85 (1972).
- de Ricqlès, A. in *A Cold Look at the Warm-Blooded Dinosaurs* (eds Thomas, R. D. K. & Olson, E. C.) Ser. 28, 103–139 (Amer. Assoc. Adv. Sci. Selected Symp., Westview Press, Boulder, 1980).
- Reid, R. E. H. The histology of dinosaurian bone, and its possible bearing on dinosaurian physiology. *Symp. Zool. Soc. Lond.* **52**, 629–663 (1984).
- Reid, R. E. H. On supposed Haversian bone from the hadrosaur *Anatosaurus*, and the nature of compact bone in dinosaurs. *J. Paleont.* **59**, 140–148 (1985).
- Chinsamy, A. Physiological implications of the bone histology of *Syntarsus rhodesiensis* (Saurischia: Theropoda). *Palaeont. Afr.* **27**, 77–82 (1990).
- Varricchio, D. J. Bone microstructure of the upper Cretaceous theropod dinosaur *Troodon formosus*. *J. Vert. Paleont.* **13**, 99–104 (1993).
- Gebhardt, W. Über funktionell wichtige Anordnungsweisen der feineren und gröberen Bauelemente des Wirbeltierknochens. II. Spezieller Teil. 1. Der Bau der Haversschen Lamellensysteme und seine funktionelle Bedeutung. *W. Roux Archiv für Entwicklungsmechanik der Organismen* **20**, 187–334 (1906).
- Amprino, R. & Godina, G. La struttura della ossa nei vertebrate. Ricerche comparative negli anfibi e negli amnioti. *Commentat. Pontif. Acad. Scient.* **11**, 329–463 (1947).
- Sereno, P. C. The evolution of dinosaurs. *Science* **284**, 2137–2147 (1999).
- de Ricqlès, A., Meunier, F. J., Castanet, J. & Francillon-Vieillot, H. in *Bone 3, Bone Matrix and Bone Specific Products* (ed. Hall, B. K.) 1–78 (CRC, Boca Raton, Florida, 1991).
- Curtis, T. A., Ashrafi, S. H. & Weber, D. F. Canalicular communication in the cortices of human long bones. *Anat. Rec.* **212**, 336–344 (1985).
- Pawlicki, R. Metabolic pathways of the fossil dinosaur bones Part III. Intermediary and other osteocytes in the systems of metabolic pathways of dinosaur bone. *Folia Histochem. Cytobiol.* **22**, 91–98 (1984).
- Gauthier, J. in *The Origin of Birds and the Evolution of Flight* (ed. K. Padian). *Mem. Calif. Acad. Sci.* **8**, 1–55 (1986).

- Holtz, T. R. The phyletic position of the Tyrannosauridae: implications for theropod systematics. *J. Paleont.* **68**, 1100–1117 (1994).
- de Ricqlès, A. Recherches paléohistologiques sur les os longs de Tétrapodes. VII. Sur la classification, la signification fonctionnelle et l'histoire des tissus osseux des Tétrapodes. Partie I. *Ann. Paléont. (Vertebres)* **61**, 51–129 (1975).
- Pawlicki, R., Korb, A. & Kubiak, H. Cells, collagen fibrils and vessels in dinosaur bone. *Nature* **211**, 655–657 (1966).
- Rimblot-Baly, E., de Ricqlès, A. & Zylberberg, L. Analyse paléohistologique d'une série de croissance partielle chez *Lapparentosaurus madagascariensis* (Jurassique moyen): essai sur la dynamique de croissance d'un dinosaure sauroptère. *Ann. Paléont. (Invert. - Vert.)* **81**, 49–86 (1995).
- Amprino, R. La structure du tissu osseux envisagée comme expression de différences dans la vitesse de l'accroissement. *Arch. Biologie* **58**, 315–330 (1947).
- Giraud-Guille, M. M. Twisted plywood architecture of collagen fibrils in human compact bone osteons. *Calcif. Tissue Int.* **42**, 167–180 (1988).
- McMahon, J. M., Boyde, A. & Bromage, T. G. Pattern of collagen fiber orientation in the ovine calcaneal shaft and its relation to locomotor-induced strain. *Anat. Rec.* **242**, 147–158 (1995).
- Raspanti, M., Guizzardi, S., Strocchi, R. & Ruggeri, A. Collagen fibril patterns in compact bone: preliminary ultrastructural observations. *Acta Anat.* **155**, 249–256 (1996).
- Castanet, J., Grandin, A., Abourachid, A. & de Ricqlès, A. Expression de la dynamique de croissance dans la structure de l'os périostique chez *Anas platyrhynchos*. *C. R. Acad. Sci. Paris* **319**, 301–308 (1996).
- Boyde, A. in *The Biochemistry and Physiology of Bone* (ed. Bourne, G.) 259–310 (Academic, New York & London, 1972).
- Currey, J. D. *The Mechanical Adaptations of Bones* (Princeton Univ. Press, Englewood Cliffs, New Jersey, 1984).
- Martin, R. B. & Burr, D. B. *Structure, Function, and Adaptation of Compact Bone* Fig. 2. 4. (Raven, New York, 1989).
- Crawford, G. N. C. The evolution of the Haversian pattern in bone. *J. Anatomy* **74**, 284–299 (1940).
- Xu, X., Tang, Z.-L. & Wang, Z.-L. A therizinosauroid dinosaur with integumentary structures from China. *Nature* **399**, 350–354 (1999).

Acknowledgements

We thank J. Ostrom for identifying two ornithomimid specimens; S. Herring, A. de Ricqlès, M. Smith and N. Wolf for assistance with the manuscript; and B. Evans, S. McCallum, D. McDougall, S. Ott-Ralph and D. Sherrard for helpful discussion. A. Busbey, P. Currie, G. Erickson, J. Joslin, A. Kemp, O. Rieflin, S. Rohwer, H.-D. Sues, C. Wood and the Woodland Park Zoo kindly made specimens available. B. Witte collected and prepared dinosaur specimens; S. Andres, D. Bennett, G. Bergsma, V. Carr, C. Chihara, J. Cyrus, M. Dyakoff, J. Haag, H. Heller, M. Hoffman, N. Huston, D. Jacobson, K. Jensen, J. Johnston, J. Kim, K. Krigbaum, T. Lee, B. Leu, M. Mielkey, B. Moorthy, J. Morton, E. Moye, H. Mull, J. Nguyen, B. Olsen, C. Pardo, L. Rende, O. Rieflin, L. Ruhwedel, T. Schommer, M. Shields, A. Smith, B. Summers, T. Ting, H. Wada, J. Weber and H. Witt ground thin sections or helped in other aspects of the study. We are grateful to the Burke Museum, the Hayashibara Museum, the University of Washington College of Arts and Sciences, and Departments of Geological Sciences and Orthodontics for support.

Correspondence and requests for materials should be addressed to J.M.R. (e-mail: rensb@u.washington.edu).

Unexpectedly similar rates of nucleotide substitution found in male and female hominids

Hacho B. Bohossian, Helen Skaletsky & David C. Page

Howard Hughes Medical Institute, Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

In 1947, it was suggested that, in humans, the mutation rate is dramatically higher in the male germ line than in the female germ line¹. This hypothesis has been supported by the observation that, among primates, Y-linked genes evolved more rapidly than homologous X-linked genes^{2–6}. Based on these evolutionary studies, the ratio (α_m) of male to female mutation rates in primates was estimated to be about 5. However, selection could have skewed sequence evolution in introns and exons^{7–10}. In addition, some of the X–Y gene pairs studied lie within chromosomal regions with substantially divergent nucleotide sequences^{7,11,12}. Here we directly compare human X and Y sequences within a large

Table 1 Numbers of nucleotide substitutions and α_m : direct estimates

	Numbers of substitutions				α_m (95% confidence interval)
	Total	On Y chromosome	On X chromosome	Unresolved	
All substitutions	441	238	175	28	1.66 (1.19–2.45)
Transitions	305	164	121	21	1.65 (1.11–2.66)
Transversions	136	74	54	7	1.68 (0.95–3.68)

region with no known genes. Here the two chromosomes are 99% identical, and X–Y divergence began only three or four million years ago, during hominid evolution^{13–15}. In apes, homologous sequences exist only on the X chromosome. We sequenced and compared 38.6 kb of this region from human X, human Y, chimpanzee X and gorilla X chromosomes. We calculated α_m to be 1.7 (95% confidence interval 1.15–2.87), significantly lower than previous estimates in primates. We infer that, in humans and their immediate ancestors, male and female mutation rates were far more similar than previously supposed.

Li *et al.* have suggested that the most accurate and precise estimates of α_m should emerge from comparison of lengthy DNA segments that are non-functional but highly similar in sequence⁷. We therefore focused upon a large region of 99% nucleotide identity between the long arm of the human X chromosome (Xq) and the short arm of the human Y chromosome (Yp). These sequences are present on human Yp because of a massive X-to-Y transposition that occurred about three or four million years ago, after divergence of the human and chimpanzee lineages^{13–15}. This Xq–Yp region is poor in genes (T. Kawaguchi *et al.*, unpublished results). The 1% divergence between the human X- and Y-linked sequences presumably reflects the random accumulation of new mutations on both the X and Y chromosomes during the last three to four million years of hominid evolution. Given this evolutionary history and the paucity of genes, the Xq–Yp region offers a nearly ideal substrate for estimation of α_m in hominids.

From within this Xq–Yp region we selected a 38.6-kb segment for detailed study. Human X and Y-chromosomal bacterial artificial chromosomes (BACs) containing this segment had been isolated in our laboratory and sequenced at the Whitehead Institute/MIT Center for Genome Research. This segment has a total content of guanine and cytosine (G+C content) of 35%; Alus, LINES, and other interspersed repeat elements account for 61% of the total sequence on both X and Y. We found no known or electronically predicted genes within this segment or within 100 kb to either side of the segment.

We compared the sequences of this 38.6-kb segment as found on the human X, human Y, chimpanzee X and gorilla X chromosomes. For the chimpanzee and gorilla X chromosomes, we generated sequencing templates by polymerase chain reaction (PCR) amplification using female genomic DNAs as starting material. As controls, we re-sequenced the corresponding portions of the human X and Y BACs, again using PCR-generated templates. In this manner, we were able to assemble 38.6 kb of virtually continuous sequence from all four chromosomes. Assembly and sequencing of genomic PCR products across the entire 38.6-kb segment were straightforward because the region's interspersed repeat ele-

ments, though numerous, were ancient and thus did not impede selection of locus-specific oligonucleotide primers. This characteristic of the region, together with the absence of genes, had led us to select it for study.

We discovered that, within this segment, the human X and Y chromosome differed at 441 nucleotides (Table 1); the two human chromosomes were 98.86% identical, in good agreement with previous estimates for the larger Xq–Yp region¹⁵. Of these 441 nucleotide substitutions, we were able to infer in 413 cases whether the mutation had occurred on the hominid X chromosome (175 cases) or on the hominid Y chromosome (238 cases). This was done by examining, for each substitution, the corresponding nucleotide position on the chimpanzee and gorilla X chromosomes. For example, if a T nucleotide was present on the human Y chromosome, but an A was present at the corresponding site on human, chimpanzee, and gorilla X chromosomes, we inferred that the primitive or ancestral state was A, and that an A-to-T substitution had occurred on the hominid Y chromosome. Conversely, if the human Y, gorilla X, and chimpanzee X chromosomes were identical to each other but differed from the human X chromosome at a particular nucleotide, we inferred that a mutation had arisen on the hominid X chromosome. We were unable to infer the chromosomal origin of the substitution at only 28 nucleotide sites; in most such cases, we observed nucleotide differences between chimpanzee and gorilla. (We also traced the chromosomal origins of small insertions and deletions—15 on the hominid X chromosome, 23 on the hominid Y chromosome—but these events were too few to merit detailed analysis.)

Using the inferred numbers of nucleotide substitutions on the hominid X and Y chromosomes, and ignoring the 28 unresolved differences, we estimated α_m by Miyata's formula:

$$Y/X = 3\alpha_m / (2 + \alpha_m)$$

where Y/X is the ratio of mutation rates on the two sex chromosomes². This formula is based on the expectation that, in any generation, two-thirds of X chromosomes are transmitted through the female germ line, the remaining one-third being transmitted through the male germ line. By contrast, all Y chromosomes are transmitted through the male germ line. Our observed Y/X ratio of $(238/175) = 1.36$ (95% confidence interval 1.12–1.65) implies that $\alpha_m = 1.66$ (95% confidence interval 1.19–2.45). Restricting the analysis to either transitions or transversions does not alter the estimate of α_m (Table 1).

Our calculations of Y/X and α_m are based on direct comparison of human X and human Y chromosomal sequences with a readily inferred ancestral sequence. All previous estimates of Y/X (and thus of α_m) in primates involved X–Y gene pairs that were too diverged to allow accurate reconstruction of ancestral sequences. Previous estimates of Y/X required construction and comparison of two trees of evolutionary distances: one tree for orthologous Y-linked

Table 2 Evolutionary distances (substitutions per 100 sites)

	Human Y	Human X	Chimpanzee X
Human X	1.18 ± 0.06		
Chimpanzee X	1.71 ± 0.07	1.56 ± 0.07	
Gorilla X	1.82 ± 0.07	1.66 ± 0.07	1.86 ± 0.07

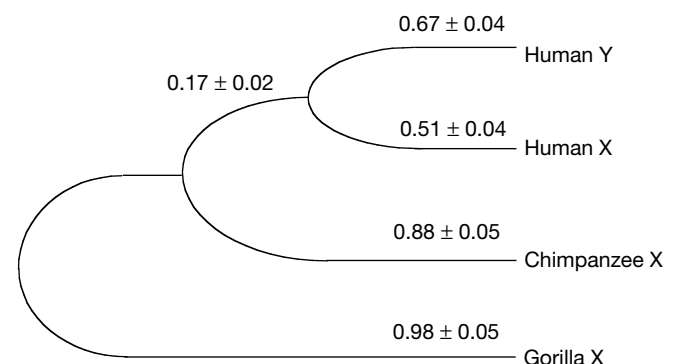


Figure 1 Phylogenetic tree of nucleotide sequences. Branch lengths were estimated from the pairwise evolutionary distances (substitutions per 100 sites) in Table 2.

sequences in several species and a second tree for orthologous X-linked sequences in the same species^{3–6}. To ensure that discrepancies between past and present findings were not attributable to different methods of calculation, we re-analysed our sequence data using the traditional method. We first estimated evolutionary distances between the human Y, human X, chimpanzee X and gorilla X sequences using the method of ref. 16 (Table 2). As expected, the resulting phylogenetic tree (Fig. 1) casts chimpanzee and gorilla as outgroups to the human X and Y sequences. The tree yields a Y/X value of $(0.67/0.51) = 1.31$ (95% confidence interval 1.05–1.55), implying that $\alpha_m = 1.55$ (95% confidence interval 1.08–2.14). Thus, recalculating Y/X and α_m by the traditional phylogenetic method yields essentially the same results as our direct calculation. The phylogenetic tree also suggests that X–Y gene conversion was not a major factor in the evolution of these sequences in hominids. (Even if some X–Y gene conversion had occurred, it would not affect our estimates of α_m .) Finally, the phylogenetic tree corroborates the conclusion¹⁵ that the X-to-Y transposition occurred within one to two million years after the divergence (roughly four to six million years ago) of the human and chimpanzee lineages.

Although our direct calculation indicated that $\alpha_m \approx 1.66$, this may be a slight underestimate. Our calculation assumed that the 1.1% divergence observed between the human X- and Y-linked sequences was entirely attributable to mutations that arose after X-to-Y transposition. However, the 38.6-kb sequence analysed may have been polymorphic (on the X chromosome) at the time of transposition, three to four million years ago. Such ancient polymorphism could account for part of the observed X–Y divergence and could result in our underestimating both Y/X and α_m . Nonetheless, recent studies of sequence diversity on X chromosomes in modern humans and chimpanzees suggest that the correction for ancient polymorphism should be modest. Assuming that X-linked sequence diversity in ancient hominids approximated that in modern human populations (4×10^{-4} per base pair, bp)¹⁷, our estimate of α_m is essentially unchanged. Alternatively, if X-linked sequence diversity in ancient hominids was as high as that in modern chimpanzee populations (1.3×10^{-3} per bp)¹⁸, then our estimate of α_m should be corrected to about 1.8 (95% confidence interval 1.15–2.87).

In any case, our estimate of α_m in hominids is much lower than previous estimates in primates (Table 3). Several factors may account for this discrepancy. Our study addressed the last three to four million years of hominid evolution. By contrast, previous experimental designs required comparisons among divergent primate lineages and thus yielded estimates of α_m averaged across broad swaths of primate evolution^{3–6}. These studies could not have detected whether α_m was lower in hominids than in other primates. Second, our study examined a much larger DNA segment and yielded more precise estimates of α_m . Large standard errors in previous estimates (Table 3) may explain, in part, the apparent disparity with present findings. Third, as has been pointed out, nucleotide sequence context may influence or bias substitution

rates^{7,11,12}. This may have affected previous estimates of α_m , as these were based on relative substitution rates in substantially diverged portions of the X and Y chromosomes^{3–6}. Such contextual bias should be negligible in our present examination of X and Y-chromosomal regions whose DNA sequences are 99% identical. Finally, and perhaps most importantly, the present analysis involved sequences which appear to be physically distant from any gene; selective neutrality can reasonably be assumed. By contrast, previous estimates of α_m were based on analyses of introns or exons. The higher estimates of α_m in past studies could reflect: (1) relaxed selective constraints on Y-linked introns and exons as compared with their X-linked homologues^{7–9}; (2) diminished mutation rates in X-linked genes^{10,19}, or both. Future studies of genes within the region of 99% X–Y identity may allow investigators to document the effects, if any, of mutation and selection bias on estimates of α_m .

Our findings suggest that substitution rates were only modestly higher in males than females during the last three to four million years of hominid evolution. If these inferences extend to modern humans, as seems likely, then they have ramifications for medical genetics. Many individuals are afflicted by autosomal dominant or X-linked recessive disorders because of new mutations that appeared in their parents' or grandparents' germlines^{1,20,21}. In several such disorders, and especially those caused by recurrence of specific substitutions at particular nucleotide positions in one gene (such as achondroplasia²²), nearly all new mutations arise in the male germline. Bolstered by studies of X–Y gene evolution in primates^{2–6}, this dramatic sex bias at a small number of extraordinarily mutable nucleotides has been taken as evidence that substitution rates across the human genome are much higher in males than in females^{20,21}. Our results suggest a re-interpretation of these medically ascertained hot spots for mutation. Our estimate of $\alpha_m \approx 1.7$ is based on the complete, diverse set of germline substitutions that accumulated within a large, selectively neutral region. Our data may provide the best global estimate to date of substitutional sex ratios in the human genome. From this point of reference, the large sex biases at some medically important hot spots appear as marked departures from global norms, underscoring the importance of unexplored interactions between sequence context and sex at these unusual sites.

Our findings also challenge the model that human mutation rates are directly proportional to the number of cell divisions, regardless of sex. Beginning with Haldane¹, high α_m values have been attributed to the much greater number of germline cell divisions in males (with mitotically active spermatogonial stem cells) than in females (where germ cells cease dividing during foetal development)^{2,3}. Our results, however, suggest that sexual asymmetry in substitution rates is far less striking than sexual asymmetry in numbers of cell divisions, at least in hominids. We suggest two possibilities for further investigation. Perhaps errors in mitotic DNA replication and repair account for a minority of germline substitutions in human genes. Alternatively, perhaps DNA replication and repair are unusually accurate in spermatogonial stem cells and their prospermatogonial precursors, which account for most of the excess cell divisions in the male germ line. □

Methods

Reference DNA sequences

We studied a 38.6-kb X–Y homologous segment which corresponds both to nucleotides 28,842–67,422 of a human X-chromosomal BAC (GenBank AC002488) and to nucleotides 88,369–127,049 of a human Y-chromosomal BAC (GenBank AC002509). These BACs, isolated in our laboratory from the California Institute of Technology A (CTA) library²³, had been sequenced at the Whitehead Institute/MIT Center for Genome Research. Interspersed repetitive elements within the 38.6-kb segment were identified electronically using RepeatMasker (<http://ftp.genome.washington.edu/RM/RepeatMasker.html>); all repeats were included in the analysis of mutations. Electronic searches employing GenScan²⁴, Grail²⁵ and BLAST²⁶ failed to identify any genes or exons within the 38.6-kb segment. Electronic searches also failed to identify any genes or exons in three adjoining Y-chromosomal BACs (GenBank AC012078, AC010094 and AC010737), all sequenced at the Washington University Genome Sequencing Centre.

Table 3 Comparison of Y/X-based estimates of α_m in primates

Genes studied	Exon or intron	DNA segment length (kb)	Y/X ratio	α_m (95% confidence interval)	Reference
ZFX/ZFY	intron	0.9	2.27	6.3 (2.6–32)	3,5
ZFX/ZFY	exon	1.2	2.0	4.2 (not reported)	4
SMCX/SMCY	intron	1.4	2.03	4.2 (2.2–10)	5
AMELX/AMELY	intron	1.1	2.16	5.14 (2.4–17)	6
Pooled data for ZFX/ZFY, SMCX/SMCY & AMELX/AMELY	introns	see above	2.15	5.06 (3.24–8.8)	6
Gene-free Xq–Yp region	neither	38.6	1.36	1.7 (1.15–2.87)	present study

Resequencing

A series of overlapping fragments, each about 1 kb in length, that collectively spanned the 38.6-kb region was generated by PCR using each of four DNAs as starting material: the human X-chromosomal BAC, the human Y-chromosomal BAC, chimpanzee female genomic DNA, and gorilla female genomic DNA. PCR primers were selected using Primer3 (ref. 27) and are available upon request. PCR products were purified on Sephadryl-S300 columns and sequenced using fluorescent-dye-terminator cycle sequencing protocols (ThermoSequenase kit; Amersham). Primers used in PCR generation of sequencing templates were also used as sequencing primers; additional sequencing primers were selected at sites internal to the PCR-generated templates.

The sequences of each of the four chromosomes (human X, human Y, chimpanzee X, gorilla X) were assembled using Sequencher 3.1 (Gene Codes Corp.). The four chromosomes were aligned using MegAlign (DNASTAR, Inc.), and the alignment was edited manually (alignment available upon request). There was only one discrepancy (T→C at nucleotide 56,776 in X-chromosomal sequence) between our PCR-generated human X and Y-chromosomal sequences and the corresponding, GenBank-deposited reference sequences.

Statistical calculations

In this direct method, α_m was calculated directly from the inferred numbers of Y and X substitutions via the formula $Y/X = 3\alpha_m/(2 + \alpha_m)$ (ref. 2). We calculated confidence intervals for ratios of substitution rates using the formula for relative risk²⁸.

We also analysed our sequence data using the method of ref. 3. We began by calculating, for each pairwise sequence comparison, the number of substitutions per 100 nucleotides¹⁶. From there we estimated branch lengths and their variances²⁹, and these values in turn enabled us to estimate Y/X and α_m (ref. 3).

When correcting estimates of α_m for ancient polymorphism, we calculated means and variances for the numbers of substitutions after X-to-Y transposition, and then calculated means and standard deviations for Y/X using the delta method³⁰.

GenBank accession numbers

Gorilla: AF190869, AF190870 and AF190871. Chimpanzee: AF190865, AF190866, AF190867 and AF190868.

Received 19 April; accepted 2 June 2000.

- Haldane, J. B. S. The mutation rate of the gene for haemophilia, and its segregation ratios in males and females. *Ann. Eugen.* **13**, 262–271 (1947).
- Miyata, T., Hayashida, H., Kuma, K., Mitsuyasu, K. & Yasunaga, T. Male-driven molecular evolution: A model and nucleotide sequence analysis. *Cold Spring Harbor Symp. Quant. Biol.* **52**, 863–867 (1987).
- Shimmin, L. C., Chang, B. H. -J. & Li, W. -H. Male-driven evolution of DNA sequences. *Nature* **362**, 745–747 (1993).
- Shimmin, L. C., Chang, B. H. -J. & Li, W. -H. Contrasting rates of nucleotide substitution in the X-linked and Y-linked zinc finger genes. *J. Mol. Evol.* **39**, 569–578 (1994).
- Chang, B. H. -J., Hewett-Emmett, D. & Li, W. -H. Male-to-female ratios of mutation rate in higher primates estimated from intron sequences. *Zool. Stud.* **35**, 36–48 (1996).
- Huang, W., Chang, B. H. -J., Gu, X., Hewett-Emmett, D. & Li, W. -H. Sex differences in mutation rate in higher primates estimated from AMG intron sequences. *J. Mol. Evol.* **44**, 463–465 (1997).
- Shimmin, L. C., Chang, B. H. -J., Hewett-Emmett, D. & Li, W. -H. Potential problems in estimating the male-to-female mutation rate ratio from DNA sequence data. *J. Mol. Evol.* **37**, 160–166 (1993).
- Charlesworth, B. The effect of background selection against deleterious alleles on weakly selected, linked variants. *Genet. Res.* **63**, 213–227 (1994).
- Li, W. -H. *Molecular Evolution* (Sinauer, Sunderland, 1997).
- McVean, G. T. & Hurst, L. D. Evidence for a selectively favourable reduction in the mutation rate of the X chromosome. *Nature* **386**, 388–392 (1997).
- Bulmer, M. Neighboring base effects on substitution rates in pseudogenes. *Mol. Biol. Evol.* **3**, 322–329 (1986).
- Wolfe, K. H., Sharp, P. M. & Li, W. -H. Mutation rates differ among regions of the mammalian genome. *Nature* **337**, 283–285 (1989).
- Page, D. C., Harper, M. E., Love, J. & Botstein, D. Occurrence of a transposition from the X-chromosome long arm to the Y-chromosome short arm during human evolution. *Nature* **311**, 119–123 (1984).
- Mumm, S., Molini, B., Terrell, J., Srivastava, A. & Schlessinger, D. Evolutionary features of the 4-Mb Xq21.3 XY homology region revealed by a map at 60-kb resolution. *Genome Res.* **7**, 307–314 (1997).
- Schwartz, A. *et al.* Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination. *Hum. Mol. Genet.* **7**, 1–11 (1998).
- Tajima, F. & Nei, M. Estimation of evolutionary distance between nucleotide sequences. *Mol. Biol. Evol.* **1**, 269–285 (1984).
- Kaessmann, H., Heissig, F., von Haeseler, A. & Paabo, S. DNA sequence variation in a non-coding region of low recombination on the human X chromosome. *Nature Genet.* **22**, 78–81 (1999).
- Kaessmann, H., Wiebe, V. & Paabo, S. Extensive nuclear DNA sequence diversity among chimpanzees. *Science* **286**, 1159–1162 (1999).
- Ellegren, H. & Fridolfsson, A. K. Male-driven evolution of DNA sequences in birds. *Nature Genet.* **17**, 182–184 (1997).
- Vogel, F. & Motulsky, A. G. *Human Genetics* (Springer, Berlin, 1997).
- Crow, J. The high spontaneous mutation rate: Is it a health risk? *Proc. Natl Acad. Sci. USA* **94**, 8380–8386 (1997).
- Wilkin, D. J. *et al.* Mutations in fibroblast growth-factor receptor 3 in sporadic cases of achondroplasia occur exclusively on the paternally derived chromosome. *Am. J. Hum. Genet.* **63**, 711–716 (1998).
- Shizuya, H. B. *et al.* Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl Acad. Sci. USA* **89**, 8794–8797 (1992).
- Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).

- Urbacher, E. C. & Mural, R. J. Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc. Natl Acad. Sci. USA* **88**, 11261–11265 (1991).
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
- Rozen, S. & Skaletsky, H. in *Bioinformatics Methods and Protocols* (eds Misener, S. & Krawetz, S. A.) (Humana, Totowa, 1999).
- Agresti, A. *Categorical Data Analysis* (Wiley, New York, 1990).
- Li, W. -H. A statistical test of phylogenies estimated from sequence data. *Mol. Biol. Evol.* **6**, 424–435 (1989).
- Bishop, Y. V. V., Feinberg, S. E. & Holland, P. W. *Discrete Multivariate Analysis* (MIT Press, Cambridge, Massachusetts, 1975).

Acknowledgements

We thank A. Schwartz for identifying homologous X and Y-chromosomal BACs, colleagues at the Whitehead Institute/MIT Centre for Genome Research for sequencing those BACs, and J. Bradley, A. Chakravarti, B. Charlesworth, A. Clark, D. Haig, T. Kawaguchi, L. Kruglyak, F. Lewitter, Y.-F. Lim, D. Reich, W. Rice, S. Rozen, C. Tilford and J. Wang for comments on the manuscript. Supported in part by the NIH.

Correspondence and requests for materials should be addressed to D.C.P. (e-mail: dcp@wi.mit.edu).

Evolvability of an RNA virus is determined by its mutational neighbourhood

Christina L. Burch & Lin Chao

Department of Biology, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0116, USA

The ubiquity of mechanisms that generate genetic variation has spurred arguments that evolvability, the ability to generate adaptive variation, has itself evolved in response to natural selection^{1,2}. The high mutation rate of RNA viruses is postulated to be an adaptation for evolvability^{3,4}, but the paradox is that whereas some RNA viruses evolve at high rates^{4,5}, others are highly stable^{5,6}. Here we show that evolvability in the RNA bacteriophage $\phi 6$ is also determined by the accessibility of advantageous genotypes within the mutational neighbourhood (the set of mutants one or a few mutational steps away). We found that two $\phi 6$ populations that were derived from a single ancestral phage repeatedly evolved at different rates and toward different fitness maxima. Fitness measurements of individual phages showed that the fitness distribution of mutants differed between the two populations. Whereas population A, which evolved toward a higher maximum, had a distribution that contained many advantageous mutants, population B, which evolved toward a lower maximum, had a distribution that contained only deleterious mutants. We interpret these distributions to measure the fitness effects of genotypes that are mutationally available to the two populations. Thus, the / evolvability of $\phi 6$ is constrained by the distribution of its mutational neighbours, despite the fact that this phage has the characteristic high mutation rate of RNA viruses.

Populations of $\phi 6$ carrying a deleterious mutation have been shown to recover fitness more often by different compensatory mutations than by a back mutation that reverted the deleterious mutation⁷. The mutations were compensatory because their benefit was conditional (epistatic) on the presence of the deleterious mutation in the same viral genome. Thus, the recovered populations could have evolved new co-adapted gene complexes⁸ whose subcomponents, the epistatic and the deleterious mutations, were potentially harmful when alone. Such complexes are effectively different evolutionary solutions to maximizing fitness in our culture

conditions, and they correspond to different fitness maxima (adaptive peaks) in the viral adaptive landscape (the projection of fitness over the sequence space)^{9–11}. Because different peaks can also differ in magnitude, the adaptive landscape could impose an additional constraint on the evolvability of the organism.

To determine whether the new complexes had different maxima, we tested their evolvabilities by propagating two recovered populations (hereafter A and B) and allowing them to evolve for 100 generations. Whereas population A continued to evolve higher fitness, population B did not (Fig. 1). The mean fitness of population B remained unchanged for 100 generations, and the final fitness of population A was higher than the final fitness of population B ($P < 0.05$, d.f. = 4; two-tailed t -test). Although these results suggest the possibility of different fitness maxima, they might also have been the result of chance alone. Although both populations A and B had the same initial fitness (Fig. 1), the quicker rise in population A might have been the result of an advantageous mutation appearing first in that population. Given more time, population B might also rise to the same level of fitness as population A.

To test whether chance or different evolvabilities best explained our results, we isolated a single clone at the 50-generations time point ($t = 50$; Fig. 1) from both population A and population B, and propagated each clone (hereafter clones A and B) for 100 generations, which subjected the clones to additional evolution. A single clone was isolated from each population to control for the chance appearance of an advantageous mutation in population A. The cloning resets to zero the level of genetic variation and forces clones A and B to start evolving with the same level of variation. Thus, if the difference were due to the chance appearance of an advantageous mutation, the evolutionary trajectory of clones A and B should be identical; however, if the difference were due to different evolvabilities, clones A and B would be expected to retrace the original trajectories of populations A and B.

Clones A and B were each propagated with a fivefold replication and the resulting evolutionary trajectories were unexpected (Fig. 2). After 100 generations, clone A evolved an increase in mean fitness, whereas clone B evolved a loss. The five clone A replicates exhibited a mean fitness gain of $\Delta W = +0.132$ (significantly greater than zero; $P < 0.0145$; d.f. = 4; two-tailed one sample t -test), whereas the five clone B replicates exhibited a mean loss in fitness of $\Delta W = -0.248$ (significantly less than zero; $P < 0.0001$; d.f. = 4; same test), where $\Delta W = W_{100} - W_0$, and W_0 and W_{100} are fitness at $t = 0$ and $t = 100$ generations. We could interpret the behaviour of clone A as the result of natural selection, but the evolution of clone B into a population of lower fitness was puzzling. However, a careful inspection of the data showed that clones A and B, which had been randomly chosen, were by chance outlying members of

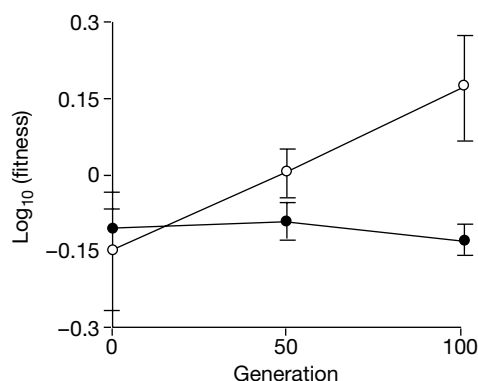


Figure 1 Fitness trajectories of populations A (open circles) and B (filled circles) successively propagated using a bottleneck of 1,000 phage. Each point is mean fitness $\pm$ standard errors ($\log_{10}$ -transformed; $n = 6$) for the population. Fitness measured by the standard assay.

populations A and B. Whereas the mean fitness of populations A and B were 1.01 and 0.81 at the time that the clones were isolated ($t = 50$ generations; Fig. 1), the fitness of clones A and B were 0.71 and 1.05 before evolution ($t = 0$; Fig. 2). Thus, a possible interpretation of the behaviour of clones A and B is that both populations A and B contained a wide distribution of genotypes with different fitness values. Because the clones came from the tails of the distribution, the populations that they generated evolved back to match the mean of their respective original populations (A and B), by regenerating through mutations the other members of the distribution. Because clone B belonged to the upper tail, many of its mutant offspring were less fit and clone B evolved towards a lower mean fitness.

Key to this interpretation is that clone B generated less fit or deleterious mutations, and that clone A generated more fit or advantageous mutations as the two clones were allowed to evolve. Because frozen samples of the evolving populations were stored, testing this interpretation required only measuring the fitness distribution of individual genotypes in the samples. Because a large number of samples was needed, however, we could not use our standard fitness assay as it is too laborious. We therefore resorted to using plaque size as a correlate of fitness. Both assays allow the virus to grow and form plaques on a lawn of its bacterial host, but, whereas our standard assay titred the actual number of virus within a plaque, the plaque-size assay measured the area of the plaque. A calibration had shown that $\log_{10}(\text{fitness})$ by our standard assay and plaque size were highly correlated and linearly related within the range of our measurements ($R^2 = 0.87$, $P < 0.0001$, d.f. = 8; data not shown).

Plaque-size measurements of the fitness distribution of individual clones within the populations descending from clones A and B strongly support the interpretation that deleterious mutations were responsible for the fitness drop of clone B (Fig. 3). The distribution was determined for the time points $t = 0$, 50 and 100 generations during the evolution of one replicate of clone A and one of clone B (from Fig. 2). At $t = 0$, both clones A and B have not accumulated mutations and the distribution reflects the variation owing to

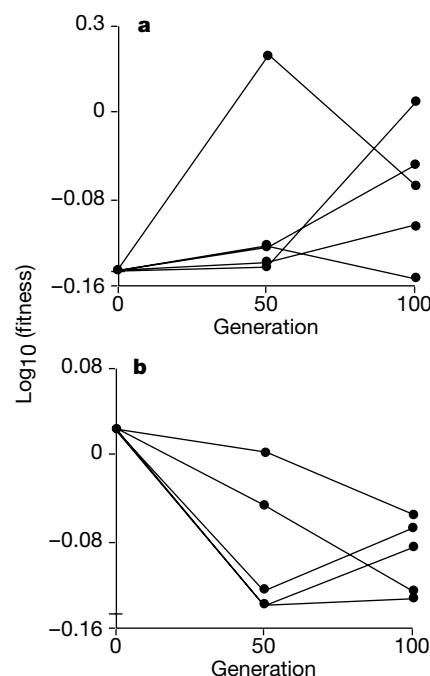


Figure 2 Fitness trajectories of five replicate populations founded by clone A (a) and clone B (b). Each point is the mean fitness of one replicate population ($\log_{10}$ -transformed; $n = 3$). Fitness measured by the standard assay.

measurement error. The mean plaque size at $t = 0$ therefore provides the estimate of plaque size for clones A and B before evolution. A visual inspection of the distribution data at $t = 50$ and 100 shows clearly that larger plaques accumulated in the lineage descending from clone A, but not in the lineage descending from clone B. Whereas 85% of the plaques at $t = 100$ were greater than the mean plaque size at $t = 0$ for clone A, 75% were smaller than the $t = 0$ mean for clone B. A comparison of mean plaque size at $t = 0$ and 100 for each clone confirms this apparent difference. Mean plaque size was larger by 1.13 mm^2 for clone A ($P < 0.0001$; d.f. = 78; two-tailed t -test) and smaller by 0.43 mm^2 for clone B ($P < 0.001$, d.f. = 78; same test). Thus, most of the mutants generated by clone A are advantageous mutations and most of those generated by clone B are deleterious mutations.

Because our isolate of clone B had a higher fitness than the mean of population B, our test of the evolvability of populations A and B was inadvertently strengthened. As indicated above, our initial expectation was that different evolvabilities would have caused the clones to retrace the original trajectories of populations A and B. We anticipated that clone B would have the same fitness as the mean of population B; however, if clone B had simply retraced the evolution of population B, different evolvabilities would have been concluded, but the scope of the conclusion would have been limited. We would not have been able to determine whether the difference was due to populations A and B evolving to adaptive peaks of different magnitudes or evolving to the same peak but along shoulders of different inclines. However, because clone B was better than average and it evolved to a lower fitness, a fitness valley separates populations A and B. Thus, the different evolvabilities of populations A and B resulted from their proximity to different adaptive peaks that were also of different magnitudes (see Fig. 4 for a graphical interpretation). Although this result does not challenge the possibility that evolvability is evolved, it shows that evolvability is constrained by more than just the ability to generate genetic variation.

Our finding of different evolvabilities for clones A and B parallels the results of a study¹² on the evolution of resistance in HIV treated

with protease inhibitor. In that study, resistance was found to be costly, but the virus was able to evolve compensatory mutations to reduce the cost. The most surprising outcome, however, was that the fitness of the resistant/compensatory virus evolved to be higher than the fitness of the ancestral virus, which was sensitive to the inhibitor. But, although this important result suggests that the ancestral and the resistant/compensatory virus are on different adaptive peaks, it is only suggestive because the ancestral virus was not allowed to evolve for the same number of generations as the resistant/compensatory virus. Thus, a control is lacking. The contribution of our present study is to provide evidence for multiple peaks by using clones A and B as controls for each other.

Given the high mutation rate of RNA viruses, the wide fitness distribution of genotypes in populations A and B is not unexpected. In the case of population B, our results suggest that the distribution is an equilibrium between a high deleterious mutation rate and selection against the deleterious mutations. In population genetics, the equilibrium is called the mutation–selection balance, and the fitness reduction due to the deleterious mutations is the mutational load. In virology, the equilibrium is the quasispecies of Eigen^{10,13}. The quasispecies is generally assumed to be of evolutionary and clinical importance in RNA viruses because it provides the genetic variation needed to respond to natural selection. Nonetheless, evidence for the quasispecies has been primarily based on reports of high polymorphism levels, rapid adaptation rates and the frequent appearance of escape mutants in these viruses⁴. Such evidence is inadequate because it demonstrates the potential benefit rather than the existence of quasispecies. The evolution of clone B in our studies provides the first demonstration of a quasispecies in an RNA virus. By showing that a high-fitness member of the distribution evolves to a lower mean fitness as the distribution of deleterious mutants is regenerated, our results confirm a previous prediction¹⁴ for quasispecies.

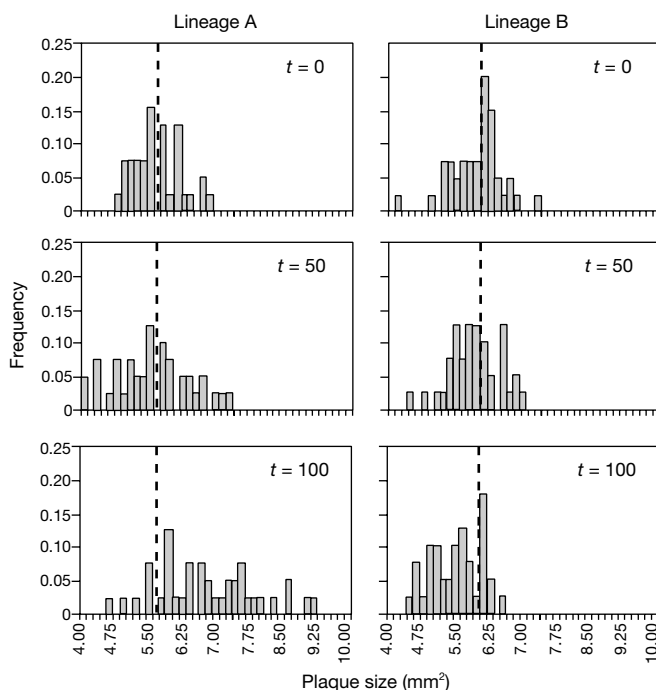


Figure 3 Plaque-size distribution over time for one clone A replicate and one clone B replicate (both from Fig. 2). Dashed line represents mean plaque size at $t = 0$ and is drawn on plots for $t = 50$ and $t = 100$ generations as a reference to highlight changes in the distribution over time.

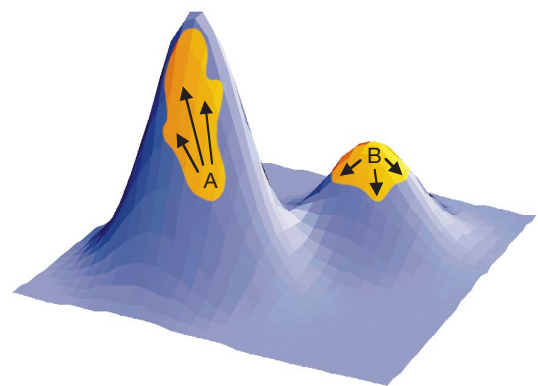


Figure 4 An interpretative representation of the evolution of clones A and B on a two-peaked adaptive landscape. The x and y planes represent sequence space on a horizontal surface. Because all possible genotypes are imagined to be projected onto the surface, where neighbouring genotypes differ by one mutation, the scale of sequence space is units of genetic distance^{9,10}. Increasing fitness is represented vertically on the z axis. The resulting plot of the fitness of individual genotypes generates the fitness surface, which is coloured blue. The letters A and B represent the fitness values of clones A and B (generation 0 in Fig. 2). The yellow patches represent the fitness values of genotypes in the populations derived from clones A and B after 100 generations of evolution. The arrows represent the dominant direction of evolution on the fitness surface. Because clone A is a distance from the higher fitness peak, it evolves into a population of higher fitness. In contrast, clone B already occupies the lower fitness peak and all mutations are deleterious. Because of the high mutation rate of $\phi 6$, selection is unable to remove all of the deleterious mutations. Thus, clone B evolves into a population of lower fitness. The evolution of lowered fitness by clone B implies that a fitness valley separates populations A and B. The size and shape of the blue and yellow surfaces are chosen arbitrarily and presented solely for illustration.

Two cautionary points, one perhaps encouraging and another less so, emerge from our study. First, it is evident that a high mutation rate is not always beneficial. The adaptive neighbourhood of an adaptive peak can constrain evolvability, despite a high mutation rate. A high rate can lead to a high mutational load, as seen during the evolution by clone B. Thus, evolutionary changes in a population of RNA viruses, whether under clinical treatment or otherwise, should not automatically be interpreted as adaptive to the virus. Second, as had already been noted¹², if multiple peaks exist, a shift to a new adaptive peak can be triggered by a costly resistance mutation and the ensuing compensatory mutations. If some peaks are higher, drugs should be used with caution not only because resistance can evolve but also because drug treatment can lead to the evolution of viruses that are more fit rather than less. □

Methods

Strains and culture conditions

The RNA bacteriophage $\phi 6$ used in this study is a laboratory clone descended from the original isolate¹⁵. Populations A and B were generated in a previous study⁷. They are the end points of populations propagated using bottlenecks of 100 and 33 phage, respectively, to allow fitness recovery after acquisition of a common deleterious mutation. *Pseudomonas syringae* sv. *phaseolicola*, the standard host of $\phi 6$, was obtained from the American Type Culture Collection (ATCC# 21781), and *Pseudomonas pseudocaligenes* ERA, an alternate host, was obtained from L. Mindich. Details of diluting, filtering, culture and storage of phage and bacteria are published^{16,17}. All phage and bacteria were grown in LC medium at 25 °C.

Phage propagation

Details for protocol have been described⁷. Briefly, phage were plated onto a lawn of standard host *P. phaseolicola* and incubated to allow the phage to reproduce and form plaques on the lawn. After 24 h, a number of plaques were randomly chosen from the plate, phage were collected from the plaques, and then plated on a fresh lawn to start a new growth cycle. One growth cycle on the lawn corresponds to about five generations, and cycles can be repeated for as long as desired. The number of plaques chosen each cycle was 1,000 for all of the populations in this study. This number renders the effective population size of all of the populations to be equivalent by enforcing the same size of population bottleneck. A bottleneck size of 1,000 is sufficiently large to promote evolution by natural selection in $\phi 6$ (ref. 7). Viral samples from all time points were stored frozen for later fitness assays.

Because the goal of the study was to examine the evolvability of these populations, and not their co-evolvability with their bacterial host, bacterial evolution was prevented by filtering out the bacteria after each growth cycle. Thus, every growth cycle was initiated with a fresh lawn made from a frozen (non-evolving) sample of the bacteria.

Standard fitness assay

The protocol we used has been described¹⁸. A test phage and a reference phage were mixed at a ratio of roughly 1:1 and plated on a *P. phaseolicola* lawn at a density of 400 phage per plate. After a 24-h incubation, the resulting plaques were collected. Fitness was measured as $W = R_1/R_0$, where R_0 and R_1 are, respectively, the ratio of the phage (test to reference) before and after growth on the lawn. This fitness assay measures the realized growth rate of the test phage relative to the reference. A value of $W = 1.0$ or $\log_{10}(W) = 0.0$ indicates equal fitness. To differentiate the two phage, the reference phage was marked with a host range mutation that allows growth on a new host, the bacterium *P. pseudocaligenes*.

Plaque-size determination

Plaque sizes were determined by plating phage from a population on a lawn of *P. phaseolicola* and incubating for 24 h. Phage were plated at a low density (< 50 phage per plate) to ensure non-overlapping plaques. Plaques growing on the lawn, denoted test plaques, could have been used to quantify the fitness of individual viruses in the population; however, the size of a single test plaque is very sensitive to environmental factors and does not provide a precise measurement of fitness. To increase the precision, each test plaque was cut from the lawn, suspended in broth medium, diluted and plated on a scoring plate at a density of roughly 60 plaques per plate. Pictures were taken of the scoring plates and the size of plaques on each plate was determined using Scion Image Version 3b (Scion Corporation, Frederick, MD). The average plaque size on each scoring plate was then used as the measure of plaque size for the corresponding test plaque. Each variate in the distributions presented in Fig. 3 is such an average for a single test plaque.

Statistics

All analyses were done as described¹⁹.

Received 22 February; accepted 15 May 2000.

1. Radman, M., Matic, I. & Taddei, F. Evolution of Evolvability. *Ann. N. Y. Acad. Sci.* **870**, 146–155 (1999).

2. Wagner, G. P., Chiu, C. & Hansen, T. F. Is Hsp90 a regulator of evolvability? *J. Exp. Zool.* **285**, 116–118 (1999).
3. Holland, J. et al. Rapid evolution of RNA genomes. *Science* **215**, 1577–1585 (1982).
4. Domingo, E. & Holland, J. J. in *The Evolutionary Biology of Viruses* (ed. Morse, S. S.) 161–184 (Raven, New York, 1994).
5. Fraile, A. et al. A century of tobamovirus evolution in an Australian population of *Nicotiana glauca*. *J. Virol.* **71**, 8316–8320 (1997).
6. Chao, L. Evolution of sex and the molecular clock in RNA viruses. *Gene* **205**, 301–308 (1997).
7. Burch, C. L. & Chao, L. Evolution by small steps and rugged landscapes in the RNA virus $\phi 6$. *Genetics* **151**, 921–927 (1999).
8. Fenster, C. B., Galloway, L. F. & Chao, L. Epistasis and its consequences for the evolution of natural populations. *Trends Ecol. Evol.* **12**, 282–286 (1997).
9. Wright, S. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. Sixth Int. Congress Genetics* **1**, 356–366 (1932).
10. Eigen, M. New concepts for dealing with the evolution of nucleic acids. *Cold Spring Harbor Symp. Quant. Biol.* **52**, 307–320 (1987).
11. Kauffman, S. A. *The Origins of Order*. (Oxford Univ. Press, New York, 1993).
12. Nijhuis, M. et al. Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. *AIDS* **13**, 2349–2359 (1999).
13. Eigen, M. On the nature of viral quasispecies. *Trends Microbiol.* **4**, 212–214 (1996).
14. Schuster, P. & Swetina, J. Stationary mutant distributions and evolutionary optimization. *Bull. Math. Biol.* **50**, 636–660 (1988).
15. Vidaver, A. K., Kolski, R. K. & van Etten, J. L. Bacteriophage $\phi 6$: a lipid-containing virus of *Pseudomonas phaseolicola*. *J. Virol.* **11**, 799–805 (1973).
16. Mindich, L., Sinclair, J. F., Levine, D. & Cohen, J. Genetic studies of temperature-sensitive and nonsense mutants of bacteriophage $\phi 6$. *Virology* **75**, 218–223 (1976).
17. Chao, L., Trang, T. T. & Trang, T. T. The advantage of sex in the RNA virus $\phi 6$. *Evolution* **46**, 289–299 (1997).
18. Chao, L. Fitness of RNA virus decreased by Muller's ratchet. *Nature* **348**, 454–455 (1990).
19. Sokal, R. R. & Rohlf, F. J. *Biometry* 3rd edn (Freeman, San Francisco, 1995).

Acknowledgements

We thank our laboratory group for feedback, D. Metzgar for comments on manuscript and L. Mindich for discussion and stocks. Support was provided to C.L.B. by a predoctoral fellowship from the Howard Hughes Medical Institute and a Dissertation Improvement Grant from the National Science Foundation. Support was provided to L.C. by the University of California San Diego and the National Institutes of Health (NIGMS).

Correspondence and request for materials should be addressed to C.L.B. (e-mail: cburch@biomail.ucsd.edu).

Behaviourally driven gene expression reveals song nuclei in hummingbird brain

Erich D. Jarvis^{*}, Sidarta Ribeiro[†], Maria Luisa da Silva[‡], Dora Ventura[§], Jacques Viellard^{||} & Claudio V. Mello[‡]

^{*} Department of Neurobiology, Duke University Medical Center, Box 3209, Durham, North Carolina 27710, USA

[†] Laboratory of Animal Behavior, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

[‡] PG Neurociências e Comportamento, USP, Sao Paulo, SP, Brazil

^{||} Depto. de Zoologia, UNICAMP, CP 6109 13083-970 Campinas, SP, Brazil

[§] Depto. de Psicologia Experimental, USP, São Paulo, SP, Brazil

Hummingbirds have developed a wealth of intriguing features, such as backwards flight, ultraviolet vision, extremely high metabolic rates, nocturnal hibernation, high brain-to-body size ratio and a remarkable species-specific diversity of vocalizations^{1–4}. Like humans, they have also developed the rare trait of vocal learning, this being the ability to acquire vocalizations through imitation rather than instinct^{5,6}. Here we show, using behaviourally driven gene expression in freely ranging tropical animals, that the forebrain of hummingbirds contains seven discrete structures that are active during singing, providing the first anatomical and functional demonstration of vocal nuclei in hummingbirds. These structures are strikingly similar to seven forebrain regions that are involved in vocal learning and production in songbirds and

parrots^{7–13}—the only other avian orders known to be vocal learners⁵. This similarity is surprising, as songbirds, parrots and hummingbirds are thought to have evolved vocal learning and associated brain structures independently^{5,14}, and it indicates that strong constraints may influence the evolution of forebrain vocal nuclei.

We conducted our study at a Nature Reserve of the Museu de Biologia Mello Leitão (Espírito Santo, Brazil), an enclave of the Atlantic Tropical Forest with over 30 hummingbird species¹⁵. We focused on the sombre hummingbird (*Aphantochroa cirrhochloris*) and the rufous-breasted hermit (*Glaucis hirsuta*) (Fig. 1). We identified brain areas involved in vocal communication, by monitoring expression of the transcriptional regulator ZENK (an acronym for Zif-268, Egr-1, NGFI-A and Krox-24) in freely ranging birds after hearing and vocalizing behaviours. ZENK messenger RNA synthesis in the brain is driven by neuronal depolarization, and its detection can be used to identify select regions that are activated by specific stimuli or behaviours¹⁶. This methodology has allowed us to map vocal communication areas throughout the brains of songbirds^{12,17} and a parrot¹³ (in both laboratory and natural¹⁸ settings) without disrupting the natural behaviour of the birds. This is important as hummingbird singing behaviour can be difficult to obtain under captivity, and it is currently not possible to identify relevant brain areas in freely ranging small animals using other methods such as electrophysiology.

Hummingbirds often sing while perched on a tree branch, and feed on nectar from nearby flowers between singing bouts. We located tree perches where individual birds sang most frequently, and set feeding traps consisting of a sugar-water bottle (a flower substitute) inside a small cage on nearby tree branches. The birds were caught upon entering the cage at a specific time after a certain behavioural state. We compared three groups: silent controls (birds caught early in the morning before the start of the dawn chorus); hearing only (birds caught around the same time after hearing a 25–30-min conspecific song playback, but that did not sing in response); hearing and vocalizing (birds caught early in the morning after hearing song and singing one or more song bouts

per min during the 25–30-min period) $n = 3$ males per group for *Aphantochroa* and one male per group for *Glaucis*. The birds were killed immediately upon capture and their brains processed for ZENK mRNA expression^{12,17}.

Relative to silent controls, hearing only birds showed hearing-induced ZENK expression in seven brain areas that are conserved among avian species^{13,17}: five telencephalic (NCM, CMHV, PC, Ndc, Ai); one thalamic (DIP); and one mesencephalic (MLd) (Figs 2 and 3a; for all abbreviations, see Box 1). Relative to hearing only birds, hearing and vocalizing birds showed vocalizing-induced ZENK expression in eight discrete areas. Of these, seven are telencephalic, which we named and placed into three groups: first, a 'posterior-medial cluster' containing two nuclei (VMN and VMH); second, an 'anterior cluster' containing three nuclei (VAH, VAN and VAP); and last, a 'posterior-lateral cluster' containing two nuclei (VLN and VA) (Figs 2 and 3a). The most marked vocalizing-induced expression occurred in VAN and VLN (Fig. 2b). The eighth structure was DM in the mesencephalon (Figs 2a and 3a), a conserved avian vocal nucleus^{13,19}. ZENK expression in these eight structures was proportional to the number of song bouts produced during the 25–30-min singing period (Fig. 3b).

All seven forebrain structures with vocalizing-induced expression have distinct histological features that differentiate them from surrounding tissues (Fig. 2d). The same seven structures were found in Nissl-stained sections from males of two other hummingbird species that we caught at the Museu Mello Leitão: swallow-tailed (*Eupetomena macroura*) and white-throated (*Leucochloris albicollis*). The four species that we studied cover the two hummingbird lineages, *Glaucis* being one of the most ancient Phaethornithinae (hermits), *Aphantochroa* and *Eupetomena* two ancient Trochilinae (non-hermits), and *Leucochloris* a more recently derived Trochilinae (ref. 1; and K. L. Schuchmann and R. Bleiweiss, personal communication). Thus, forebrain vocal nuclei appear to have been

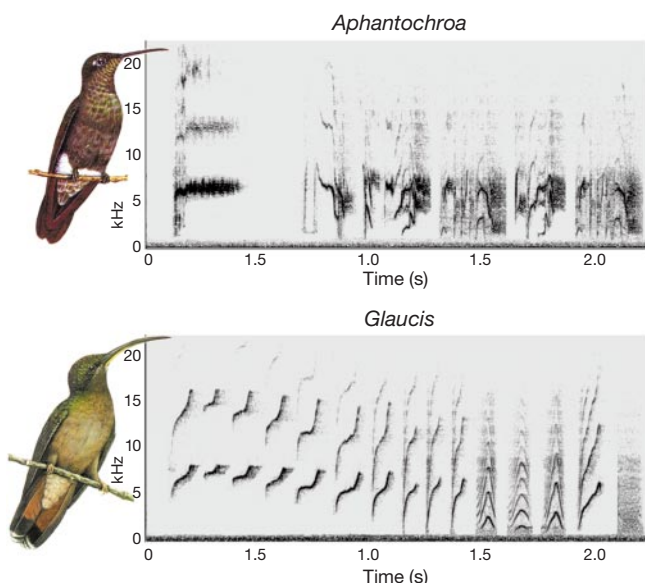


Figure 1 Song sonograms (frequency versus time) of the hummingbird species studied: *Aphantochroa cirrhochloris* and *Glaucis hirsuta*¹⁵. *Aphantochroa* has a stereotyped song, consisting of an introductory note followed by several renditions of a two-note sequence: individual notes have complex frequency modulations. *Glaucis* sings a more varied whistle-like song consisting of several introductory notes (not shown) followed by note sequences of ascending and descending frequencies that can be produced in different orders (A. Ferreira *et al.*, personal communication).

Box 1

Anatomical structures

A, archistriatum; AAC, central nucleus of the anterior archistriatum; AC, anterior commissure; ACM, caudomedial archistriatum; Ai, intermediate archistriatum; Area X, area X of the paleostriatum; Av, nucleus avalanche; Cb, cerebellum; cp, choroid plexus; CMHV, caudomedial hyperstriatum ventrale; DIP, dorsointermediate nucleus of the posterior thalamus; DLM, medial nucleus of the dorsolateral thalamus; DM, dorsomedial nucleus; DMm, magnocellular nucleus of the dorsomedial thalamus; ex, extensions of LPOm; HA, hyperstriatum accessorium; HD, hyperstriatum dorsale; Hp, hippocampus; HV, hyperstriatum ventrale; HVC, high vocal center; HVoc, complex including the oval nucleus of the hyperstriatum ventrale and surrounds; ICo, intercollicular nucleus of the mesencephalon; L2, primary telencephalic auditory area; IAHV, lateral nucleus of the anterior hyperstriatum ventrale; IAN, lateral nucleus of the anterior neostriatum; LPOm, magnocellular nucleus of the parolfactory lobe; MAN, magnocellular nucleus of the anterior neostriatum; MLd, dorsal part of the lateral mesencephalic nucleus; N, neostriatum (not the same as the mammalian neostriatum); NAoc, complex including the oval nucleus of the anterior neostriatum and surrounds; NCM, caudomedial neostriatum; Ndc, dorsocaudal neostriatum; NIDL, neostriatum intermedium pars dorsolateralis; Nlf, nucleus interfascialis; NLc, central nucleus of the lateral neostriatum; nXlts, tracheosyringeal subdivision of the hypoglossal nucleus; OC, optic chiasm; OM, occipitomesencephalic tract; OT, optic tectum; Ov, nucleus ovoidalis of the thalamus; P, paleostriatum; PC, caudal paleostriatum; PP, paleostriatum primitivum; RA, robust nucleus of the archistriatum; S, septum; T, thalamus; v, ventricle; VA, vocal nucleus of the archistriatum; VAH, vocal nucleus of the anterior hyperstriatum ventrale; VAN, vocal nucleus of the anterior neostriatum; VAP, vocal nucleus of the anterior paleostriatum; VLN, vocal nucleus of the lateral neostriatum; VMH, vocal nucleus of the medial hyperstriatum ventrale; VMN, vocal nucleus of the medial neostriatum.

present early among hummingbirds. Vocal learning has been demonstrated directly by raising birds in isolation from conspecifics for one species of the Trochilinae lineage, and indirectly through the analysis of individual variability in seven species of the Phaethornithinae lineage^{1,4–6}. Thus, it is generally assumed that vocal learning was also present early in hummingbirds.

To our knowledge, our results represent a first anatomical and functional demonstration of vocal control brain nuclei in hummingbirds. Moreover, we have identified a whole set of forebrain vocal control structures in a vocal learning order—something that has taken years in other species using different methodologies. This now provides a map for future anatomical, physiological and behavioural investigations. Interestingly, both songbirds and budgerigars (a parrot), have also been shown to have exactly seven forebrain structures with singing-induced ZENK expression^{12,13}. Three of these nuclei are at the same relative positions in the anterior telencephalon in parrots, hummingbirds and songbirds (Fig. 4, structures in red). The other four nuclei are in different locations of the posterior and/or lateral telencephalon (Fig. 4,

structures in yellow) but within the same brain subdivisions (Table 1). These nuclei have marked morphological similarities across orders. For example, parrot NLC^{11,13}, hummingbird VLN (Fig. 2) and songbird HVC⁷ bulge into the overlying ventricle. Parrot AAC^{11,13}, hummingbird VA (Fig. 2) and songbird RA⁷ have an oval shape and constitute cytoarchitectonically very distinct nuclei within the archistriatum. The posterior-lateral nuclei in songbirds and budgerigars (Fig. 4, structures in yellow) are part of a pathway whose output is to the syrinx^{11,20}, and in songbirds controls production of learned vocalizations^{7,9,12}. The anterior nuclei (Fig. 4, structures in red) control vocal learning in songbirds⁸, and are part of a pathway comparable to cortico-basal ganglia-thalamo-cortical loops^{10–12} in mammals, which participate in the learning and maintenance of sequential motor actions dependent on sensorimotor integration²¹.

Our findings have implications for the evolution of brain structures that control a complex behaviour. Vocal learning is a rare trait known to occur in only three groups of birds (parrots, hummingbirds and songbirds) and three groups of mammals (humans, cetaceans (whales/dolphins) and bats)^{5,6,14,22,23}. Because they are

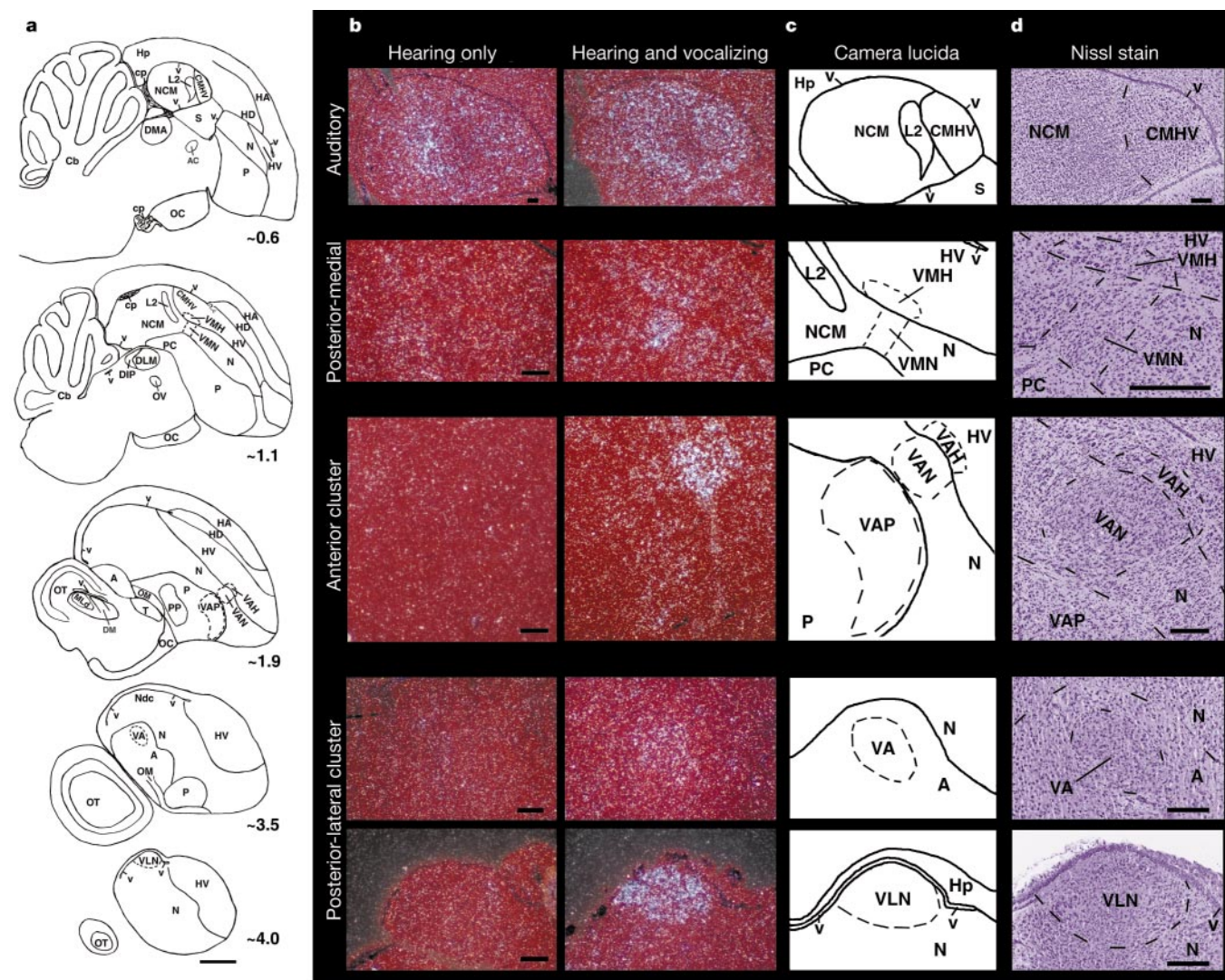


Figure 2 Identification of vocal control brain areas. **a**, Camera lucida drawings of parasagittal sections from *Aphantochroa*. Dashed lines, regions of vocalizing-induced ZENK expression. Numbers indicate mm from midline. **b**, Dark-field views of ZENK-hybridized sections. White-silver grains, ZENK expression; red background, Nissl stain. Only NCM and CMHV are shown for hearing-induced regions. **c**, Detailed drawings of brain regions from the hearing and vocalizing bird in **b**. **d**, Bright-field view of Nissl-stained reference sections containing vocalizing-induced areas, at higher magnification than in **b**

and **c** for visualization of cytoarchitectonic features: VMH contains cells larger than the surround; VMN is rectangular with cells organized into columns; VAH is small and flat; VAN is semi-round with cells larger than the surround; VAP (Nissl not shown) is crescent-shaped and contains a sub-population of large cells; VA is oval, darkly staining, and with higher cell density than the surround; VLN bulges into the overlying ventricle. Orientation: dorsal is up, anterior to the right. Scale bars, 1 mm (**a**); 0.25 mm (**b–d**). For abbreviations, see Box 1.

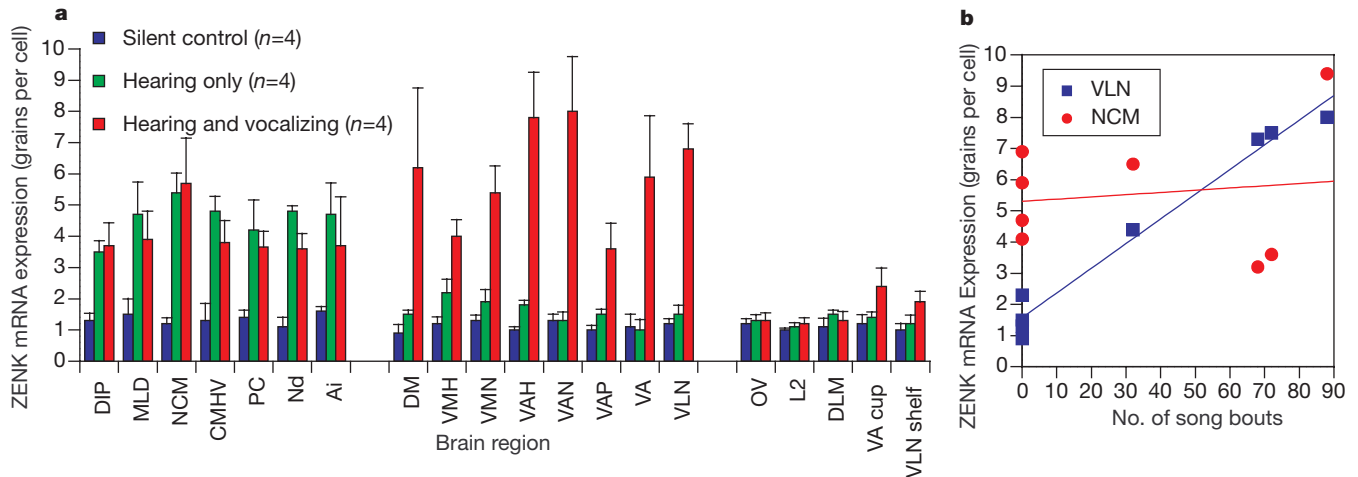


Figure 3 Quantification of ZENK expression. **a**, Regions with hearing-induced expression (left, P , 0.03 to <0.0001), vocalizing-induced expression (centre, P , 0.02 to <0.0005), or no induced expression (right, P , 0.2 to 0.84; two-tailed unpaired t -test). Of the latter, three are part of the auditory (OV, L2) or vocal (DLM) pathways, and do not show induction in songbirds and budgerigars^{12,13,17}, two (VLN* and VA*) represent regions immediately ventral to VLN and rostroventral to VA, and correspond topographically to songbird HVC shelf and RA

cup, which show hearing-induced expression in songbirds^{12,17}. Plotted are mean $\pm$ s.e.m. **b**, ZENK expression is proportional to amount of singing for structures with vocalizing-induced expression (P , 0.013 to <0.0001 ; r , 0.819–0.988) but not for those with hearing-induced expression (P , 0.10 to <0.93 ; r , 0.037–0.36; analysis of variance linear regression); VLN and NCM are representative examples. For abbreviations, see Box 1.

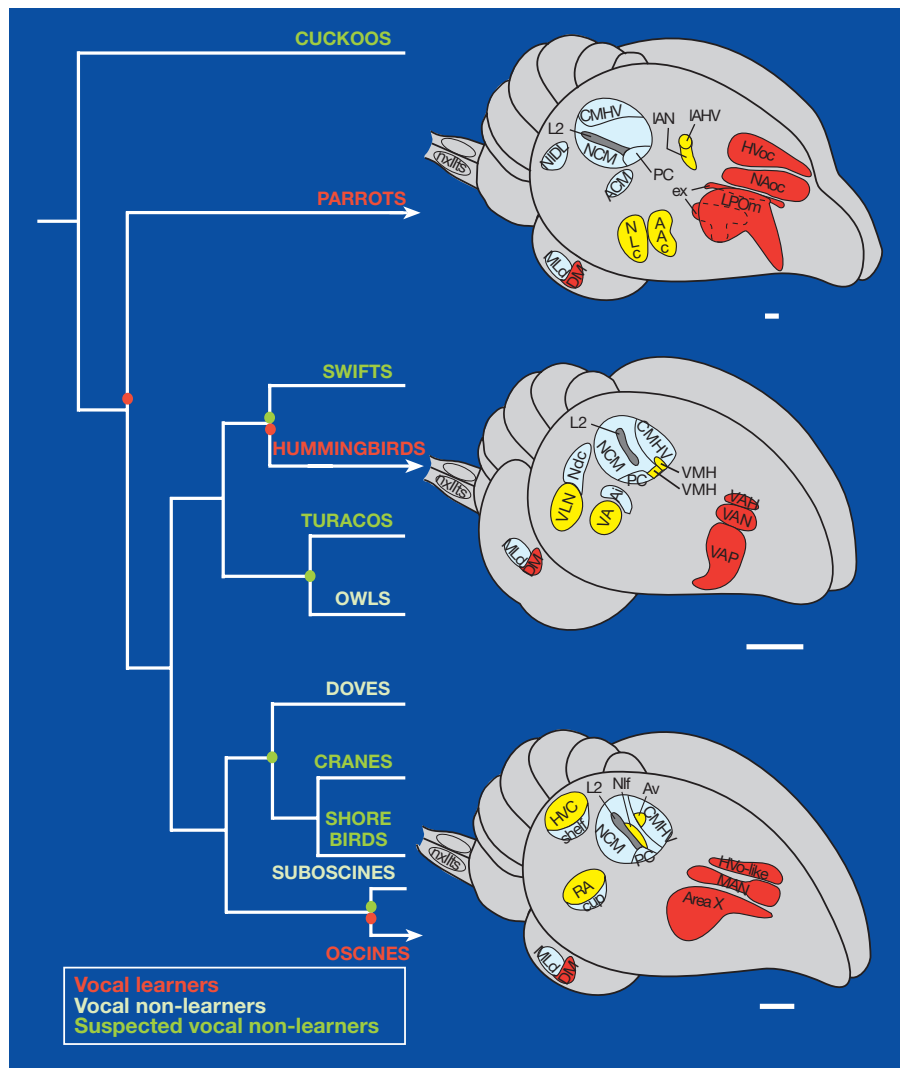


Figure 4 Comparative brain anatomy of hearing- and vocalizing-induced ZENK expression in avian vocal learners. Left, partial phylogenetic tree based on Sibley and Ahlquist²⁴, with one common species name per order; colours indicate evidence for vocal learning^{5,6,27–30}. Red dots, ancestral nodes of possible independent gain of vocal learning; green dots, ancestral nodes of possible independent loss of vocal learning. Parrots are the

oldest vocal learning order, oscine songbirds the most recent²⁴. Right, semi three-dimensional brain renditions of vocal learners. Blue, hearing-induced regions; red, similarly positioned vocalizing-induced regions; yellow, differently positioned vocalizing-induced regions^{12,13}. Scale bars, 1 mm. For abbreviations, see Box 1.

Table 1 Telencephalic subdivisions and their vocalizing-activated nuclei across vocal learning avian orders*

Brain Subdivision	Parrot	Hummingbird	Songbird
Hyperstriatum	HVoc IAHV	VAH VMH	HVo-like Av
Neostriatum	NAoc NLc IAN	VAN VLN VMN	MAN HVC Nlf
Archistriatum	AAc	VA	RA
Paleostriatum	LPOm	VAP	Area X

*Correspondences are based on relative position and cytoarchitectonic features for the three orders, in addition to connectivity data for songbirds and parrots (budgerigars)^{17,11–14,20}. For abbreviations, see Box 1.

phylogenetically separated by vocal non-learners²⁴ (Fig. 4), it is thought that the three avian vocal learning groups, and presumably the mammalian ones, evolved vocal learning independently^{5,14}. Similarly, the associated forebrain vocal control structures are absent in avian vocal non-learners, and it is believed that songbirds, parrots and presumably hummingbirds evolved such structures independently^{14,25}. Modern birds evolved from a common ancestor thought to have lived about 65 million years ago near the Cretaceous/Tertiary transition²⁶. Of the descendants, parrots are the oldest vocal learning order, followed by hummingbirds and then oscine songbirds²⁴ (Fig. 4). According to the dominant hypothesis^{5,14}, our results indicate that within the past 65 million years 3 out of 23 avian orders²⁴ may have independently evolved 7 similar forebrain vocal structures for a complex behaviour (Fig. 4). This would suggest that the evolution of these structures is under strong epigenetic constraints; in which case, similar structures may have also evolved in vocal learning mammals (humans, cetaceans and bats). Alternatively, vocal learning and associated brain structures may have been present in a common ancestor to avian vocal learners. In this regard, there is a shift in the posterior forebrain vocal structures from more anterior-lateral to posterior-medial positions, in accordance with the relative age of the vocal learning orders (Fig. 4). This hypothesis requires that the forebrain vocal structures were lost in the intervening vocal non-learning orders at least four times independently (Fig. 4). Such a loss could be due to the considerable expense required to maintain vocal learning and associated brain structures, with many birds possibly evolving in adaptive zones that did not require complex learned vocalizations. Another alternative hypothesis is that avian vocal non-learners have some rudimentary form of forebrain vocal areas that were previously missed by Nissl and tract tracing studies^{14,19,25}. If true, this would constitute a challenge to the hypothesis that forebrain vocal structures are unique to vocal learners¹⁴. □

Methods

Behaviour

The behaviour of all birds was monitored by video taping: only those birds that could be monitored during the entire observation period were captured. Video recordings were used to score number of song bouts produced by the hearing and vocalizing group during the observation period. For the hearing only group, playbacks of digitally recorded conspecific song (three song bouts per min, each 3–4 s long, from a bird of another locale) were presented from a nearby speaker (3–4 ft). To capture the birds, a feeding bottle containing 24% sucrose inside a cage was used. After 25–30 min in one of the behavioural conditions, a string attached to a stick holding the cage door open was pulled and the bird caught on a regular visit to the feeding bottle. The birds were immediately killed by decapitation, and their brains were dissected, placed in cryogenic embedding medium, and frozen in dry ice; sex was confirmed by direct inspection of the gonads.

Gene expression

Serial parasagittal (right hemisphere) and frontal (left hemisphere) frozen sections (10 µm) were cut throughout the entire brain of each bird. One section every 0.1 mm of each brain (~100 sections per brain, totalling 1,200 sections) was processed for ZENK expression by *in situ* hybridization with a ³⁵S radioactively labelled riboprobe for canary ZENK, followed by emulsion autoradiography^{12,17}. Unhybridized adjacent sections (20 µm) were stained with cresyl violet and used as reference for identification of cytoarchitectonic boundaries (Fig. 2d). Quantification (Fig. 3) was performed by counting silver grains over cells¹². Regions where ZENK expression was significantly higher in

hearing only compared with silent control animals represent areas activated by hearing conspecific song; regions where ZENK expression was higher in hearing and vocalizing compared with hearing only animals represent areas activated by singing. This strategy reveals all known telencephalic vocal control nuclei in songbirds and parrots, and identified previously undetected ones^{12,13}. No obvious differences were detected between *Aphantochroa* and *Glaucis*, and thus their values were grouped (total *n* = 4 per group).

Received 28 March; accepted 30 May 2000.

- Schuchmann, K.-L. in *Handbook of the birds of the world: barn-owls to hummingbirds* (eds del Hoyo, J., Elliott, A. & Sargatal, J.) 468–680 (Lynx Ediciones, Barcelona, 1999).
- Ventura, D. F. & Takase, E. Ultraviolet color discrimination in the hummingbird. *Invest. Ophthalmol. Vis. Sci.* **35**, 2168 (1994).
- Rehkmper, G., Schuchmann, K. L., Schleicher, A. & Zilles, K. Encephalization in hummingbirds (Trochilidae). *Brain Behav. Evol.* **37**, 85–91 (1991).
- Vieillard, J. Catálogo sonográfico dos cantos e piados dos beija-flores do Brasil. *Boletim do Museu de Biologia "Mello Leitão" Série Biologia* **58**, 1–20 (1983).
- Nottebohm, F. The origins of vocal learning. *Am. Nat.* **106**, 116–140 (1972).
- Baptista, L. F. & Schuchmann, K. L. Song learning in the anna hummingbird (*Calypate anna*). *Ethology* **84**, 15–26 (1990).
- Nottebohm, F., Stokes, T. M. & Leonard, C. M. Central control of song in the canary, *Serinus canarius*. *J. Comp. Neurol.* **165**, 457–486 (1976).
- Scharff, C. & Nottebohm, F. A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. *J. Neurosci.* **11**, 2896–2913 (1991).
- Yu, A. C. & Margoliash, D. Temporal hierarchical control of singing in birds. *Science* **273**, 1871–1875 (1996).
- Hessler, N. A. & Doupe, A. J. Singing-related neural activity in a dorsal forebrain-basal ganglia circuit of adult zebra finches. *J. Neurosci.* **19**, 10461–10481 (1999).
- Durand, S. E., Heaton, J. T., Amateau, S. K. & Brauth, S. E. Vocal control pathways through the anterior forebrain of a parrot (*Melopsittacus undulatus*). *J. Comp. Neurol.* **377**, 179–206 (1997).
- Jarvis, E. D., Scharff, C., Grossman, M. R., Ramos, J. A. & Nottebohm, F. For whom the bird sings: context-dependent gene expression. *Neuron* **21**, 775–788 (1998).
- Jarvis, E. D. & Mello, C. V. Molecular mapping of brain areas involved in parrot vocal communication. *J. Comp. Neurol.* **419**, 1–31 (2000).
- Brenowitz, E. A. Comparative approaches to the avian song system. *J. Neurobiol.* **33**, 517–531 (1997).
- Ruschi, A. *Hummingbirds of State of Espírito Santo* (Rios Editora, São Paulo, 1982).
- Chaudhuri, A. Neural activity mapping with inducible transcription factors. *Neuroreport* **8**, 5–9 (1997).
- Mello, C. V. & Clayton, D. F. Song-induced ZENK gene expression in auditory pathways of songbird brain and its relation to the song control system. *J. Neurosci.* **14**, 6652–6666 (1994).
- Jarvis, E. D., Schwabl, H., Ribeiro, S. & Mello, C. V. Brain gene regulation by territorial singing behavior in freely ranging songbirds. *Neuroreport* **8**, 2073–2077 (1997).
- Wild, J. M., Dongfeng, L. & Eagleton, C. Projections of the dorsomedial nucleus of the intercollicular complex (DM) in relation to respiratory-vocal nuclei in the brainstem of pigeon (*Columbia livia*) and zebra finch (*Taeniopygia guttata*). *J. Comp. Neurol.* **377**, 392–413 (1997).
- Nottebohm, F., Kelley, D. B. & Paton, J. A. Connections of vocal control nuclei in the canary telencephalon. *J. Comp. Neurol.* **207**, 344–357 (1982).
- Lidsky, T. I., Manetto, C. & Schneider, J. S. A consideration of sensory factors involved in motor functions of the basal ganglia. *Brain Res.* **356**, 133–146 (1985).
- Guinee, L. H. & Payne, K. B. Rhyme-like repetitions in songs of humpback whales. *Ethology* **79**, 295–306 (1988).
- Esser, K. H. Audio-vocal learning in a non-human mammal: the lesser spear-nosed bat *Phyllostomus discolor*. *Neuroreport* **5**, 1718–1720 (1994).
- Sibley, C. G. & Ahlquist, J. E. *Phylogeny and Classification of Birds: A Study in Molecular Evolution* (Yale Univ. Press, New Haven, 1990).
- Kroodsmas, D. E. & Konishi, M. A suboscine bird (eastern phoebe, *Sayornis phoebe*) develops normal song without auditory feedback. *Anim. Behav.* **42**, 477–487 (1991).
- Feduccia, A. Explosive evolution in tertiary birds and mammals. *Science* **267**, 637–638 (1995).
- Koenig, C. Vocal patterns as interspecific isolating mechanisms in screech owls of the genus *Otus* (Aves: Strigidae) of southern South America. *Stuttg. Beitr. Natkd. A. Biol.* **0(511)**, 1–35 (1994).
- Payne, R. B. in *Current Ornithology* (ed. Johnston, R. J.) 87–126 (Plenum, New York, 1986).
- Miller, E. H. in *Ecology and evolution of acoustic communication in birds* (eds Kroodsmas, D. E. & Miller, E. H.) 241–257 (Cornell Univ. Press, Ithaca, 1996).
- Ellis, D. H., Swengel, S. R., Archibald, G. W. & Kepler, C. B. A sociogram for the cranes of the world. *Behav. Processes* **43**, 125–151 (1998).

Acknowledgements

We thank the Museu de Biologia Mello Leitão (MBML) in Espírito Santo, Brazil, S. Mendes and D. Loss for providing a natural space and services that made this project possible. We also thank the MBML and A. Ruschi for permission to partially reproduce hummingbird illustrations by E. Demonte; P. Rousselot, K. S. Leon and A. Ferreira for help with recordings, sonograms and behavioural scoring; P. Delgado for histological assistance; S. Baumwell and L. Moore for help in manuscript preparation; S. Durand, R. Mooney and S. Nowicki for comments on the manuscript; L. Katz and N. Cant for use of microscope equipment; J. Ahlquist for discussions on avian evolution; C. Cunningham for assistance with phylogenetic analysis; and F. Nottebohm for his support. This project was approved by the Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); funding was provided by the Kluge Trust Fund and Duke University start-up funds to E.D.J., an NIDCD grant to C.V.M., and personal funds by E.D.J. and C.V.M. We dedicate this paper to the memory of L. Baptista, a pioneer of vocal communication in hummingbirds.

Correspondence and requests for materials should be addressed to C.V.M. (e-mail: mello@rockvax.rockefeller.edu) or E.D.J. (e-mail: jarvis@neuro.duke.edu).

NO is necessary and sufficient for egg activation at fertilization

Richard C. Kuo^{*†}, Gregory T. Baxter[‡], Stuart H. Thompson^{*†}, Stephen A. Stricker[§], Chris Patton[†], Joseph Bonaventura^{||} & David Epel[†]

^{*}Neurosciences Program, Stanford University School of Medicine, Stanford University, Stanford, California 94305, USA

[†]Department of Biological Sciences, Hopkins Marine Station, Stanford University, Pacific Grove, California 93950, USA

[‡]Cornell Nanofabrication Facility, Knight Laboratory, Cornell University, Ithaca, New York 14853, USA

[§]Department of Biology, University of New Mexico, Albuquerque, New Mexico 87131, USA

^{||}Department of Cell Biology, Duke University Medical Center, Durham, North Carolina 27710, USA

[¶]Nicholas School of the Environment, Duke Marine Biomedical Center, Pivers Island, North Carolina 28516, USA

The early steps that lead to the rise in calcium and egg activation at fertilization are unknown but of great interest—particularly with the advent of *in vitro* fertilization techniques for treating male infertility and whole-animal cloning by nuclear transfer. This calcium rise is required for egg activation and the subsequent events of development in eggs of all species^{1,2}. Injection of intact sperm or sperm extracts can activate eggs, suggesting that sperm-derived factors may be involved. Here we show that nitric oxide

synthase is present at high concentration and active in sperm after activation by the acrosome reaction. An increase in nitrosation within eggs is evident seconds after insemination and precedes the calcium pulse of fertilization. Microinjection of nitric oxide donors or recombinant nitric oxide synthase recapitulates events of egg activation, whereas prior injection of oxyhaemoglobin, a physiological nitric oxide scavenger, prevents egg activation after fertilization. We conclude that nitric oxide synthase and nitric-oxide-related bioactivity satisfy the primary criteria of an egg activator: they are present in an appropriate place, active at an appropriate time, and are necessary and sufficient for successful fertilization.

We first determined whether nitric oxide (NOS) is present and active in sea urchin gametes. Immunoblot analysis of sperm homogenates with a polyclonal antibody specific for the highly conserved³ calmodulin (CaM)-binding domain of rat neuronal NOS (nNOS) reveals a single immunoreactive band of approximate relative molecular mass 140,000 ($M_r \approx 140K$) (Fig. 1a). Pre-incubation of the anti-nNOS antibody with its peptide antigen blocks immunoprecipitation of the 140K protein (Fig. 1a), whereas immunoprecipitation with protein G-agarose alone does not precipitate the 140K protein (Fig. 1a). When homogenates are probed with a monoclonal 'universal' NOS (uNOS) antibody recognizing domains conserved among the inducible, endothelial and neuronal

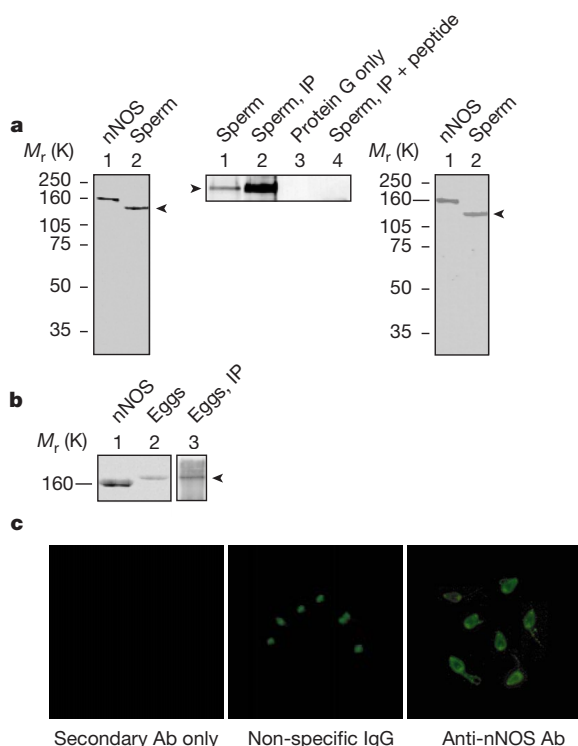


Figure 1 NOS in *S. purpuratus* gametes. **a**, Sperm homogenates (50 µg) probed with anti-nNOS antibody (left) or anti-uNOS antibody (right). Immunoprecipitation (IP) of NOS (middle). Lane 1, sperm; lane 2, IP with anti-nNOS; lane 3, protein G-agarose alone; lane 4, IP with anti-nNOS peptide antigen (25 µg). Recombinant nNOS was used at 200 ng. **b**, Egg homogenates (50 µg) probed with anti-uNOS antibody; IP of egg homogenates probed with anti-nNOS is shown in lane 3. **c**, NOS immunolabelling of sperm stained with secondary antibody (Ab) only (left), nonspecific rabbit IgG with secondary (middle), or anti-nNOS with secondary antibody (right).

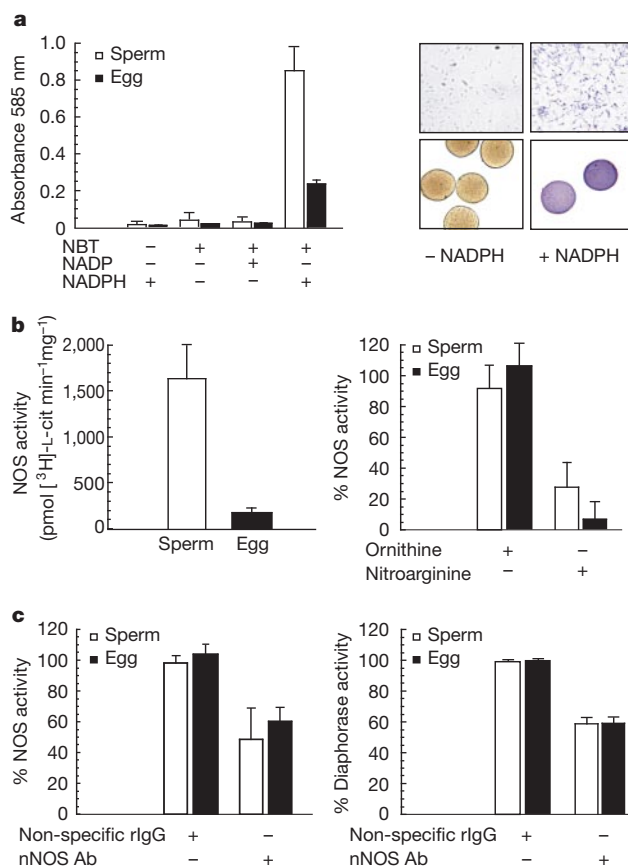


Figure 2 NOS activity in *S. purpuratus* gametes. **a**, Diaphorase activity of gamete homogenates in the presence or absence of NBT, NADP or NADPH as indicated. Absorbance at 585 nm was measured and normalized to total protein (left). NADPH diaphorase staining of gametes in the presence or absence of NADPH (right). **b**, [³H]arginine to [³H]citrulline conversion in gamete homogenates (left). Activity (relative to controls) in the presence of ornithine (1.5 mM) or N^G-nitro-L-arginine (100 µM) (right). **c**, Immunodepletion of NOS activity from gamete homogenates. NOS activity measured by citrulline conversion (left) or NADPH-dependent diaphorase activity (right) after immunodepletion with anti-nNOS antibody (Ab).

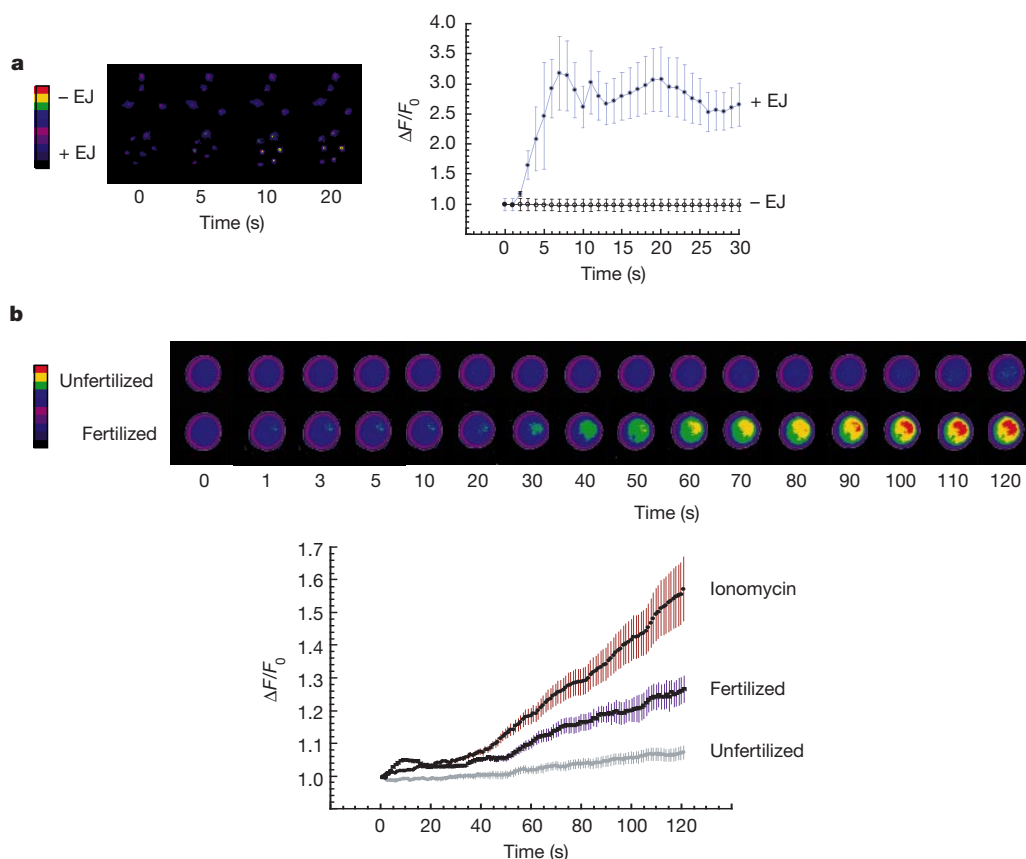


Figure 3 Nitrosation after *S. purpuratus* gamete activation. **a**, Nitrosation in sperm after acrosome reaction. Representative images (left) after treatment with seawater (top) or egg jelly (EJ) (bottom). $\Delta F/F_0$ intensity (right) for egg-jelly-stimulated sperm or seawater-treated sperm. **b**, Nitrosation in eggs at fertilization. Representative images at 0, 1, 3, 5 and 10 s intervals for unfertilized and fertilized eggs (1:10,000 sperm dilution added at

time $t = 0$ (top). $\Delta F/F_0$ intensity (over entire egg) for unfertilized eggs, fertilized eggs and ionomycin-treated eggs (2 μ M) (bottom). Results are mean $\pm$ s.d. and significant ($P < 0.01$, Student's t -test, $n = 5$ and 34) for fertilized eggs. Note that the egg centre (owing to its greater thickness) is more sensitive to changes in fluorescence signal intensity.

NOS isoforms, a distinct band of $M_r \approx 140$ K is again observed (Fig. 1a). Densitometric analysis shows that the sperm NOS isoform is present at high concentration ($\sim 0.44 \pm 0.05\%$ of total NP-40 soluble sperm protein; $n = 4$). An immunoreactive band of $M_r \approx 175$ K is also present in eggs, as shown by the anti-uNOS antibody or anti-nNOS antibody (Fig. 1b).

Immunolabelling intact sperm with the anti-nNOS antibody prominently stains the head, acrosomal region and midpiece, with

minor staining of the tail (Fig. 1c). Midpiece staining is also seen with nonspecific IgG, whereas staining of the sperm head and particularly a punctate 'dot' at the anterior, acrosomal end of the sperm is only observed with anti-nNOS antibody, indicating that NOS is concentrated in these regions and consistent with sperm-borne delivery of NOS after sperm-egg fusion.

NOS is enzymatically active in both gametes, as determined by NADPH (nicotinamide-adenine dinucleotide phosphate)-dependent diaphorase activity⁴ measured spectrophotometrically or by whole-cell diaphorase staining (Fig. 2a) and by the more specific citrulline assay⁵ (Fig. 2b). Diaphorase staining is apparent in the head and tail regions of sperm and evenly distributed throughout the entirety of eggs (Fig. 2a). The citrulline conversion exhibited by gametes might result from urea cycling⁵; however, there is no inhibition with ornithine, a urea cycle inhibitor⁶, whereas N^G -nitro-L-arginine, a specific NOS inhibitor, greatly suppresses citrulline conversion activity (Fig. 2b). The citrulline conversion and diaphorase activity is depleted by NOS immunoprecipitation from gamete

Table 1 NO production after gamete activation

Gamete	Treatment	Nitrite (pmol)	DAF fluorescence (arbitrary units)
Sperm (1 μ l)	SW, pH 8	23 $\pm$ 10	22 $\pm$ 3
	Ionomycin	75 $\pm$ 30	103 $\pm$ 13
	EJ	310 $\pm$ 15	170 $\pm$ 49
Sperm (5 μ l)	SW, pH 8	75 $\pm$ 15	42 $\pm$ 6
	Ionomycin	153 $\pm$ 15	543 $\pm$ 35
	Egg jelly	465 $\pm$ 15	1,108 $\pm$ 108
Sperm (10 μ l)	SW, pH 8	138 $\pm$ 15	70 $\pm$ 14
	Ionomycin	283 $\pm$ 15	910 $\pm$ 93
	Egg jelly	610 $\pm$ 15	1,423 $\pm$ 169
Eggs (1%, v/v)	SW, pH 8	25 $\pm$ 5	25 $\pm$ 5
	Ionomycin	64 $\pm$ 4	56 $\pm$ 10
Eggs (2%, v/v)	SW, pH 8	30 $\pm$ 7	28 $\pm$ 4
	Ionomycin	82 $\pm$ 9	76 $\pm$ 8
Eggs (5%, v/v)	SW, pH 8	33 $\pm$ 8	37 $\pm$ 5
	Ionomycin	141 $\pm$ 10	112 $\pm$ 6

NO production in *S. purpuratus* gametes was determined by measuring nitrite accumulation or DAF fluorescence. Sperm (1, 5 or 10 μ l) were incubated in seawater (SW, pH 8), 5 μ M ionomycin, or egg jelly for 5 min and exogenous NO was measured. Similarly, eggs (1%, 2% or 5% suspension) were incubated in SW or 1 μ M ionomycin for 5 min and exogenous NO was measured. Ionomycin and egg jelly values were normalized to per cent acrosome reaction after subtraction of background fluorescence. Total nitrite was determined by converting fluorescence values using nitrite standard curves. Data are from at least four experiments.

Table 2 Sperm NO production is dependent on acrosome reaction

Treatment	DAF fluorescence (%)
Egg jelly	100.0 $\pm$ 0.0
Verapamil	37.8 $\pm$ 4.0
WGA	6.4 $\pm$ 1.2
(Egg jelly + verapamil) + ionomycin	75.7 $\pm$ 9.7
(Egg jelly + WGA) + ionomycin	58.3 $\pm$ 6.1

Packed sperm (5 μ l) were incubated in the presence or absence of the acrosome reaction inhibitors verapamil¹² (100 μ M) or wheat germ agglutinin¹³ (WGA; 50 μ g ml⁻¹) for 10 min and NO was measured after addition of egg jelly or egg jelly with 5 μ M ionomycin, as indicated. Under these conditions, less than 3% of sperm were acrosome-reacted as compared with $\sim 85\%$ for controls. Data are from three separate experiments.

homogenates, further showing the specificity of the anti-nNOS antibody (Fig. 2c).

We measure production of NO by gametes by accumulation of nitrite, an auto-oxidation product of NO⁷, or by fluorescence changes after nitrosation of the NO indicator dye, diaminofluorescein (DAF)⁸ (see Supplementary Information). Qualitatively similar results are seen for exogenous NO production: activation of either male or female gametes results in NO accumulation in the surrounding medium (Table 1). For sperm, egg jelly is more effective than ionomycin in eliciting NO increases that are nonlinear with increasing sperm concentration, possibly because of auto-inhibition of NOS by NO. Inhibition of sperm motility does not affect NO production. Pre-treatment with inhibitors of the acrosome reaction^{9,10} substantially prevents NO generation, but is triggered after subsequent addition of ionomycin (Table 2). Similarly, eggs activated by ionomycin treatment generate NO, although at lower levels than sperm (Table 1). One complication may be the high thiol concentration in eggs which could scavenge endogenous NO¹¹.

A cell-permeable, diacetate form of DAF allows real-time, *in vivo* visualization of NO-derived nitrosation within sperm during the acrosome reaction (Fig. 3a), and eggs at fertilization (Fig. 3b). After egg-jelly-elicited activation, nitrosation rapidly increases in sperm (Fig. 3a), primarily in the head region, with a timing similar to the acrosome reaction (Fig. 3a). Similarly, nitrosation changes occur in eggs at fertilization: an initial transient between 0 and 5 s is followed 30–35 s later by a more gradual, sustained increase over minutes (Fig. 3b). Prior injection of oxyhaemoglobin (oxyHb), a physiological scavenger of NO¹², largely eliminates NO-derived nitrosation increases in DAF fluorescence (77.3 ± 10.4% inhibition; 8 injected eggs). These findings are consistent with work showing early increases in cGMP¹³ and cADPR¹⁴, both of which are NO dependent¹⁵.

Mimicking the NO increase by increasing intracellular NO leads to egg activation. Eggs injected with the NO donor S-nitrosoacetylpenicillamine (SNAP) or recombinant nNOS with calmodulin (CaM) activate in terms of fertilization-envelope elevation (a calcium-dependent exocytotic event) (Fig. 4a) and pronuclear centring (13 out of 13 eggs exhibiting a full fertilization envelope). Bath application of NO is ineffective (ref. 15, and unpublished data). SNAP and nNOS are only effective over a narrow concentration range (see Supplementary Information); levels above the activating concentration are ineffective, and eggs exposed to high NO concentrations are in fact refractory to activation by sperm. Activation by SNAP is dependent on the presence of the releasable NO group; N-acetylpenicillamine (the non-S-nitrosylated starting material for SNAP synthesis) and light-decomposed SNAP¹⁶ are ineffective in activating eggs (Fig. 4a). Similarly, activation of eggs by the Ca²⁺/CaM dependent nNOS isoform requires CaM; CaM alone or nNOS alone are insufficient (Fig. 4a). Egg activation does not occur when SNAP or CaM/nNOS are co-injected with oxyHb (Fig. 4a).

We assessed the necessity of NO (or its equivalents) by injecting eggs with oxyHb. Fertilization envelope elevation is dose-dependently inhibited in oxyHb-injected eggs (Fig. 4b) and other calcium-dependent indicators of egg activation, pronuclear centring and first cell division, are also lacking (0 out of 15 eggs). OxyHb-inhibited eggs can, however, elevate fertilization envelopes after insemination if subsequently treated with ionomycin or micro-injected with GTPγS or cGMP, showing that calcium stores, and more specifically the inositol trisphosphate (IP₃) (ref. 17) and cGMP¹⁸ calcium-releasing pathways, are intact and operative in the presence of oxyHb. Inhibition of calcium mobilization is specific for scavenging of NO-related bioactivity by oxyHb as egg activation is not significantly affected in eggs injected with either bovine serum albumin (a nitrosylable protein), apoHb (Hb lacking the haem moiety) or cyanometHb (Hb functionally blocked at the haem

moiety) (Fig. 4b). Accordingly, the calcium rise is substantially reduced in oxyHb-injected eggs (240 ± 30 nM resting Ca²⁺ to 1.4 ± 0.2 μM peak Ca²⁺ post-fertilization compared with 280 ± 65 nM resting Ca²⁺ to 540 ± 95 nM peak Ca²⁺ in oxyHb-injected eggs; 15 control and 14 oxyHb-injected eggs with *n* = 2), indicating that sperm enter these eggs, but fail to generate a calcium rise sufficient to elicit egg activation (Fig. 4c).

The nature of the NO-related bioactivity suggests potential mechanisms connecting sperm–egg fusion to initiation of the calcium rise at fertilization. In eggs, exogenously applied NO increases cGMP and cADPR (cyclic ADP-ribose) and elicits calcium release from ryanodine receptor (RyR)-gated stores¹⁵, but the ensuing calcium level does not result in activation. Similarly, despite early increases in cGMP¹³ and cADPR¹⁴ at fertilization, this pathway is apparently not required¹⁸. Unlike exogenous NO, endogenous NO

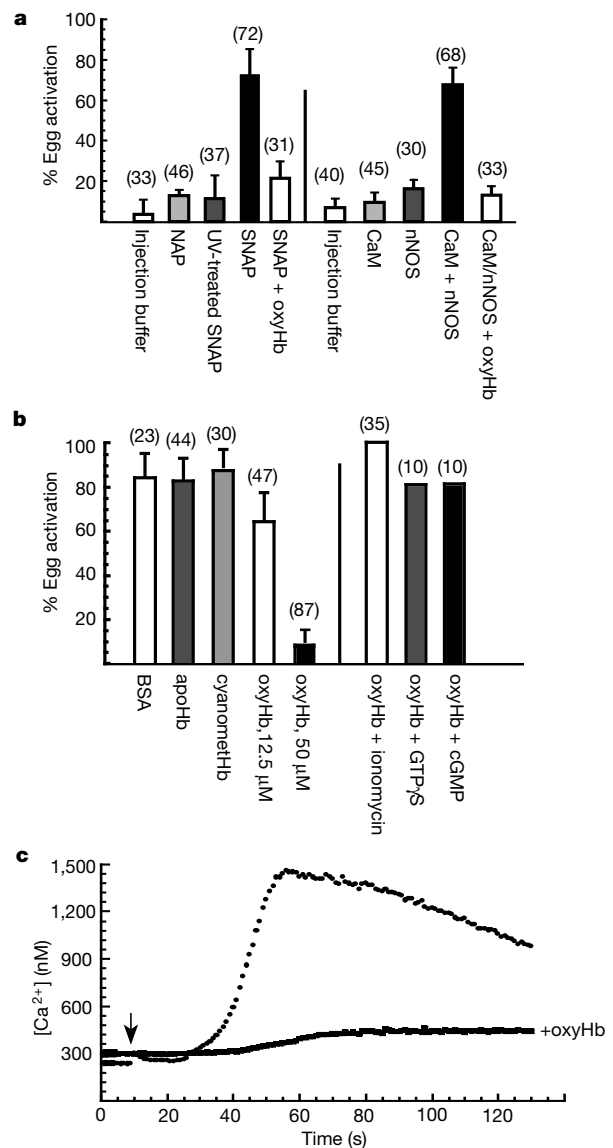


Figure 4 NO is necessary and sufficient for egg activation. **a**, Injection of the NO donor SNAP, NAP, ultraviolet-treated SNAP (~5 μM final concentration) or recombinant nNOS (~0.0075 μg final) with CaM (~0.125 μg final) or either alone. Also, co-injection of SNAP or CaM/nNOS with oxyHb (see text for details) **b**, Eggs injected with oxyHb (~12.5 μM or 50 μM protein final) or 50 μM BSA, apoHb or cyanometHb and inseminated (1:20,000 sperm dilution). Subsequent activation of oxyHb-inhibited eggs by ionomycin (2 μM) or injection of either GTPγS (~100 μM final) or cGMP (~50 μM final). Number of eggs listed in parentheses. **c**, Representative trace of calcium release in presence or absence of oxyHb measured by calcium green-2. Arrow indicates the time of sperm–egg fusion.

(or its equivalents) is sufficient for egg activation, suggesting that there are signalling pathways in addition to cGMP/cADPR that are only accessible intracellularly. One such pathway may be the direct release of calcium by S-nitrosylation of RyR stores¹⁹ which may induce additional calcium release by stimulating a calcium-sensitive phospholipase C^{21,22} and resultant IP₃ production. In addition, like phosphorylation, nitrosylation can trigger signalling intermediates such as *src*²² which in turn may elicit phospholipase C activity²³. The pleiotropic signalling provided by NO through cGMP/cADPR formation and S-nitrosylation could account for the diverse modes of calcium release seen at fertilization², and, consequently, NO may be an universal activator of eggs or oocytes. □

Methods

Materials

Gametes (*S. purpuratus* and *L. pictus*) were collected as described²⁴, and concentrated egg jelly was prepared by acid-washing eggs. For nitrite determinations, egg jelly was prepared in an artificial seawater (nASW) with reduced divalent cations comprising (in mM): 519 NaCl, 5 MgCl₂, 5 CaCl₂, 2 NaHCO₃, 10 KCl and 10 glycylglycine, pH 8. DAF was from Daiichi Pure Chemicals and recombinant nNOS and CaM were from Stratagene.

Immunoblotting and immunolabelling

Coelomocyte-free sperm or washed, de-jellied eggs were lysed in ice-cold buffer (1% NP-40 in 150 mM NaCl, 50 mM Tris, pH 8) with a protease inhibitor cocktail (PIC: bestatin, pepstatin, aprotinin, leupeptin, AEBSE, E-64 and SBTI), pelleted, and the supernatants boiled immediately before separation by SDS–poly acrylamide gel electrophoresis and immunoblotting. Egg homogenates required immunoprecipitation before immunoblotting with anti-nNOS antibody. Polyclonal anti-nNOS antibody (1:2,000) raised against the peptide sequence [T₇₂KRRAGIFKKLA₇₃₉C] corresponding to the CaM-binding domain of rat nNOS (Affinity Bioreagents), or mouse ascites anti-uNOS (1:5) raised against conserved domains of iNOS, eNOS and nNOS (gift from R. Johns) were used. Anti-nNOS was used at 1:200 dilution for immunoprecipitation. Anti-nNOS peptide antigen was provided by P. Schwartz (Affinity Bioreagents). Image Pro 4.0 was used for densitometry with recombinant nNOS for comparison. Protein was quantitated by the Bradford protein microassay (BioRad) using BSA as a standard. Anti-nNOS (1:50) and Cy2-conjugated secondary antibody (1:50) were used for NOS immunolabelling²⁵ with 0.1% Triton X-100, 1% PVP-40, 2% BSA and 5% goat serum in PBS for the blocking and washing steps.

NO/NOS assays

NADPH-dependent diaphorase activity or histochemical staining was assessed⁴ after a 24-h fixation of gametes. For activity measurements, gametes were lysed (sperm) or sonicated (eggs) in PBS with 1% NP-40 and incubated overnight in 1 mM NADPH, 0.5 mM NBT. Immunodepletion of diaphorase activity was done after size separation (>100K molecular weight cut-off (Amicon)). [³H]arginine to [³H]citrulline conversion of gamete homogenates was assessed⁵ after two cycles of freeze–thaw lysis (sperm) or Dounce homogenization (eggs) (in mM): 50 KCl, 10 NaCl, 1 EDTA, 5 DTT, 20 HEPES with PIC; pH 8. For inhibition experiments, gamete homogenates were treated with 20% (v/v) activated Dowex to remove endogenous arginine. Nitrite determinations were performed⁷ using nASW to avoid divalent cation precipitation by the stop buffer (0.5 M EDTA in 6.7 N NaOH). nASW did not affect the percentage of spontaneous or elicited acrosome reaction as judged by the ‘thumb squash’ method¹⁹. NO determination by DAF fluorescence (490/515 nm) was done after incubating gametes in seawater containing 10 μM DAF.

NO and Ca²⁺ imaging

Eggs were loaded in 10 μM DAF-DA in seawater, pH 8 (30 min at 15 °C). Sperm were loaded similarly in calcium-free artificial seawater (pH 6), but we observed a light-dependent increase in DAF fluorescence caused by nitrosylation of endogenous thiols during loading and subsequent photolytic release of NO¹⁶. This was eliminated by including 150 μM carboxy-PTIO (a cell-impermeant NO scavenger) and 0.1 mM EDTA during incubation. After loading, eggs or sperm were washed thoroughly and transferred to a temperature-controlled chamber (20 °C) on a Nikon Diaphot microscope equipped with a SIT camera. Eggs were imaged with a ×20 Fluor (N.A. = 0.75) objective and sperm with a ×100 (N.A. = 1.4) objective. ImagePro 4.0 was used for data analysis and pseudocolouring. Calcium measurements were done using calcium green-2 and calibrated using 1% Triton X-100 to determine $F_{\max}/F_{\min}$ (ref. 26).

Microinjection

L. pictus eggs were injected to ~2% egg volume²⁴ using intracellular injection media comprising (in mM): 250 glycine, 250 K⁺-gluconate, 10 HEPES, 10% PVP-40 (w/v), pH 7.0 (adjusted with acetic acid) and eggs were attached to surfaces with 10 mg ml⁻¹ protamine sulfate. SNAP and oxyHb were loaded at 4 °C in the dark and injections were performed using a red filter to prevent auto-oxidation of the oxyHb or photolysis of the SNAP¹⁶. Dialysed, chromatographically pure oxyHb, apoHb and cyanomethHb were prepared²⁷ and concentrated at 4 °C in injection media supplemented with 0.1 mM EDTA and allowed to equilibrate for 30–45 min before insemination. The fully oxygenated status

and concentration of oxyHb were determined by Soret absorbance at 415 nm. SNAP was synthesized before use²⁸. For control experiments, SNAP was decomposed by a 5-min mercury lamp exposure¹⁶. Recombinant CaM–nNOS complexes were formed by mixing the stock CaM solution at a 1:1 ratio with the appropriate nNOS dilution in injection media and kept on ice before use. Egg batches that exhibited greater than 10% spontaneous activation or less than 80–85% activation after insemination were discarded. Control and experimental injections were done within the same batch of eggs.

Received 2 February; accepted 23 May 2000.

- Epel, D. In *Handbook of Physiology* (eds. Hoffman, J. F. & Jamieson, J. J.) 859–884 (Oxford Univ. Press, New York, 1997).
- Stricker, S. A. Comparative biology of calcium signaling during fertilization and egg activation in animals. *Dev. Biol.* **211**, 157–176 (1999).
- Regulski, M. & Tully, T. Molecular and biochemical characterization of dNOS: a *Drosophila* Ca²⁺/calmodulin-dependent nitric oxide synthase. *Proc. Natl Acad. Sci. USA* **92**, 9072–9076 (1995).
- Hope, B. T., Michael, G. J., Knigge, K. M. & Vincent, S. R. Neuronal NADPH diaphorase is a nitric oxide synthase. *Proc. Natl Acad. Sci. USA* **88**, 2811–2814 (1991).
- Bredt, D. S. & Schmidt, H. H. W. In *Methods in Nitric Oxide Research* (eds Feelisch, M. & Stamler, J. S.) 249–255 (John Wiley & Sons, Chichester, 1996).
- Giraldez, R. R. & Zweier, J. L. An improved assay for measurement of nitric oxide synthase activity in biological tissues. *Anal. Biochem.* **261**, 29–35 (1998).
- Misko, T. P., Schilling, R. J., Salvemini, D., Moore, W. M. & Currie, M. G. A fluorometric assay for the measurement of nitrite in biological samples. *Anal. Biochem.* **214**, 11–16 (1993).
- Kojima, H. et al. Detection and imaging of nitric oxide with novel fluorescent indicators: diamino-fluoresceins. *Anal. Chem.* **70**, 2446–2453 (1998).
- Guerrero, A. & Darszon, A. Evidence for the activation of two different Ca²⁺ channels during the egg jelly-induced acrosome reaction of sea urchin sperm. *J. Biol. Chem.* **264**, 19593–19599 (1989).
- Podell, S. B. & Vacquier, V. D. Wheat germ agglutinin blocks the acrosome reaction in *Strongylocentrotus purpuratus* sperm by binding a 210,000-mol.wt membrane protein. *J. Cell Biol.* **99**, 1598–1604 (1984).
- Turner, E., Kleivitt, R., Hager, L. J. & Shapiro, B. M. Ovothiol, a family of redox-active mercapto-histidine compounds from marine invertebrate eggs. *Biochemistry* **26**, 4028–4036 (1987).
- Gross, S. S. & Lane, P. Physiological reactions of nitric oxide and hemoglobin: a radical rethink. *Proc. Natl Acad. Sci. USA* **96**, 9967–9969 (1999).
- Ciapa, B. & Epel, D. An early increase in cGMP follows fertilization of sea urchin eggs. *Biochem. Biophys. Res. Commun.* **223**, 633–636 (1996).
- Kuroda, R. et al. Increase of cGMP, cADP-ribose and inositol 1,4,5-trisphosphate preceding Ca²⁺ transients in fertilization of sea urchin eggs. *Zygote* (in the press).
- Willmott, N. et al. Nitric oxide-induced mobilization of intracellular calcium via the cyclic ADP-ribose signaling pathway. *J. Biol. Chem.* **271**, 3699–3705 (1996).
- Sexton, D. J., Muruganandam, A., McKenney, D. J. & Mutus, B. Visible light photochemical release of nitric oxide from S-nitrosoglutathione: potential photochemotherapeutic applications. *Photochem. Photobiol.* **59**, 463–467 (1994).
- Crossley, I., Whalley, T. & Whitaker, M. Guanosine 5'-thiotriphosphate may stimulate phosphoinositide messenger production in sea urchin eggs by a different route than the fertilizing sperm. *Cell. Reg.* **2**, 121–133 (1991).
- Lee, S. -J., Christenson, L., Martin, T. & Shen, S. S. The cyclic GMP-mediated calcium release pathway in sea urchin eggs is not required for the rise in calcium during fertilization. *Dev. Biol.* **180**, 324–335 (1996).
- Xu, L., Eu, J. P., Meissner, G. & Stamler, J. S. Activation of the cardiac calcium release channel (ryanodine receptor) by poly-S-nitrosylation. *Science* **279**, 234–237 (1997).
- Nadai, C. D., Cailliau, K., Epel, D. & Ciapa, B. Detection of phospholipase C_γ in sea urchin eggs. *Dev. Growth. Diff.* **40**, 669–676 (1998).
- Rongish, B. J., Wu, W. & Kinsey, W. H. Fertilization-induced activation of phospholipase C in the sea urchin egg. *Dev. Biol.* **215**, 147–154 (1999).
- Akhand, A. A. et al. Nitric oxide controls *src* kinase activity through a sulphydryl group modification-mediated Tyr-527-independent and Tyr-416-linked mechanism. *J. Biol. Chem.* **274**, 25821–25826 (1999).
- Giusti, A. F., Carroll, D. J., Abassi, Y. A. & Foltz, K. R. Evidence that a starfish egg *src* family tyrosine kinase associates with PLC-γ1 SH2 domains at fertilization. *Dev. Biol.* **208**, 189–199 (1999).
- Schomer, B. M. & Epel, D. The role of changes in NADPH and pH during fertilization and artificial activation of the sea urchin egg. *Dev. Biol.* **216**, 394–402 (1999).
- Schultz, J. R., Wessel, G. M. & Vacquier, V. D. The exocytosis regulatory proteins syntaxin and VAMP are shed from sea urchin sperm during the acrosome reaction. *Dev. Biol.* **191**, 80–87 (1997).
- Kao, J. P. Y. (ed.) *Practical Aspects of Measuring [Ca²⁺] with Fluorescent Indicators*. (Academic, San Diego, 1994).
- Antonini, E. & Brunori, M. *Hemoglobin and Myoglobin in their Reactions with Ligands*. (North Holland Publishing Company, Amsterdam, 1971).
- Fehsel, K., Kroncke, K. D. & Kolb-Bachofen, V. Assays for detection of nitric oxide-induced apoptosis. *Methods Enzymol.* **269**, 426–34 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank H. Tosuji for assistance with DAF, V. Lance for haemoglobin preparation, R. Lewis and H. Schulman for helpful comments, and K. Foltz and S. Feinstein for their support during R.C.K.'s stay at UCSB. RCK is a pre-doctoral fellow of the Howard Hughes Medical Institute. This work was supported by NIH grants to D.E. and J.B. and NSF grants to S.S. and S.T.

Correspondence and requests for materials should be addressed to D.E. (email: depel@stanford.edu).

YidC mediates membrane protein insertion in bacteria

James C. Samuelson*, Minyong Chen*, Fenglei Jiang*, Ines Möller†, Martin Wiedmann‡, Andreas Kuhn‡, Gregory J. Phillips§ & Ross E. Dalbey*

* Department of Chemistry, The Ohio State University, 100 West 18th Avenue, Columbus, Ohio 43210, USA

† Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Centre, 1275 York Avenue, New York, New York 10021, USA

‡ University of Hohenheim, Institute for Microbiology and Molecular Biology, Garbenstrasse 30, D-70599 Stuttgart, Germany

§ Department of Microbiology, Iowa State University, 207 Science I, Osborn Drive, Ames, Iowa 50011, USA

The basic machinery for the translocation of proteins into or across membranes is remarkably conserved from *Escherichia coli* to humans. In eukaryotes, proteins are inserted into the endoplasmic reticulum using the signal recognition particle (SRP) and the SRP receptor, as well as the integral membrane Sec61 trimeric complex (composed of alpha, beta and gamma subunits)¹. In bacteria, most proteins are inserted by a related pathway that includes the SRP homologue Ffh^{2–5}, the SRP receptor FtsY^{6,7}, and the SecYEG trimeric complex⁸, where Y and E are related to the Sec61 alpha and gamma subunits, respectively. Proteins in bacteria that exhibit no dependence on the Sec translocase were previously thought to insert into the membrane directly without the aid of a protein machinery^{9,10}. Here we show that membrane insertion of two Sec-independent proteins requires YidC. YidC is essential for *E. coli* viability and homologues are present in mitochondria and chloroplasts. Depletion of YidC also interferes with insertion of Sec-dependent membrane proteins, but it has only a minor effect on the export of secretory proteins. These results provide evidence for an additional component of the translocation machinery that is specialized for the integration of membrane proteins.

Recent studies in mitochondria, bacteria and chloroplasts suggest a new membrane protein insertion pathway that is evolutionarily related^{11–15}. In mitochondria, a subset of proteins is sorted into the inner membrane from the matrix by a conservative pathway that seems analogous to that found in bacteria¹⁶. However, one important difference in mitochondria is that the Sec proteins are absent¹⁷. Instead, Oxa1p is required for insertion of some inner membrane proteins from the mitochondrial matrix^{11,12}. Oxa1p homologues are present in the thylakoid membrane of chloroplasts (Albino3 in *Arabidopsis thaliana*¹³) and the inner membrane of bacteria (YidC in *Escherichia coli*¹⁴).

To evaluate whether YidC is involved in membrane insertion in *E. coli*, we constructed JS7131, a YidC-depletion strain where YidC expression is under the control of the *araBAD* promoter and operator (see Methods for strain construction). The genetic composition of JS7131 as shown in Fig. 1a was confirmed by polymerase chain reaction (PCR) analysis (data not shown). YidC is essential for cell viability, as determined by streaking JS7131 cells onto Luria-Bertani (LB) agar plates containing either arabinose or glucose (Fig. 1b). In the presence of arabinose, JS7131 grows at a slightly reduced rate compared with parent strains JS71 and MC1060. In the presence of glucose, there is no growth of JS7131 even after two days at 37 °C. Moreover, a significant growth defect is observed when JS7131 cells grown overnight with arabinose are back-diluted into either minimal or rich media supplemented with glucose and grown for several generations. To confirm whether YidC is strictly regulated by arabinose, immunoblot analysis was performed (Fig. 1c). This analysis reveals that YidC expression in strain JS7131 is

induced with arabinose (+ lanes) and tightly repressed in the presence of glucose (– lanes). The carbon source does not affect the expression of YidC in the parent strain MC1060. However, in strain JS71, where two functional *yidC* genes are present, addition of arabinose results in a twofold increase of YidC. The addition of arabinose to MC1060 containing *yidC* on the multi-copy pING vector induces overexpression of YidC.

We next examined whether YidC is required for protein export by examining the kinetics of signal peptide processing of pro-OmpA (outer membrane protein A), pre-MBP (maltose binding protein), and pre-β-lactamase in cells that are depleted of YidC (Fig. 2a–c). Cells grown for 3.5 h in arabinose or glucose were pulse-labelled with [³⁵S]methionine for 20 s and chased for the indicated times. Signal peptide processing of pro-OmpA is unaffected in cells depleted of YidC (glucose lanes), indicating normal export of this protein. Although there is a slight kinetic effect on the export of MBP and β-lactamase, over 90% is translocated in each case after a 2-min chase. The membrane translocation of pre-β-lactamase is also affected by depletion of the SRP homologue Ffh and the SRP receptor FtsY, like the insertion of membrane proteins^{2,6}.

The membrane insertion of a Sec-independent protein was analysed to provide evidence for a functional relationship between YidC and Oxa1p. M13 Procoat serves as the paradigm for a protein that does not rely on the Sec translocase for membrane insertion. Signal peptide processing of Procoat is nearly non-existent in YidC-depleted cells (Fig. 2d). Using a protease accessibility procedure¹⁷, we verified that the lack of processing was indeed due to a block of

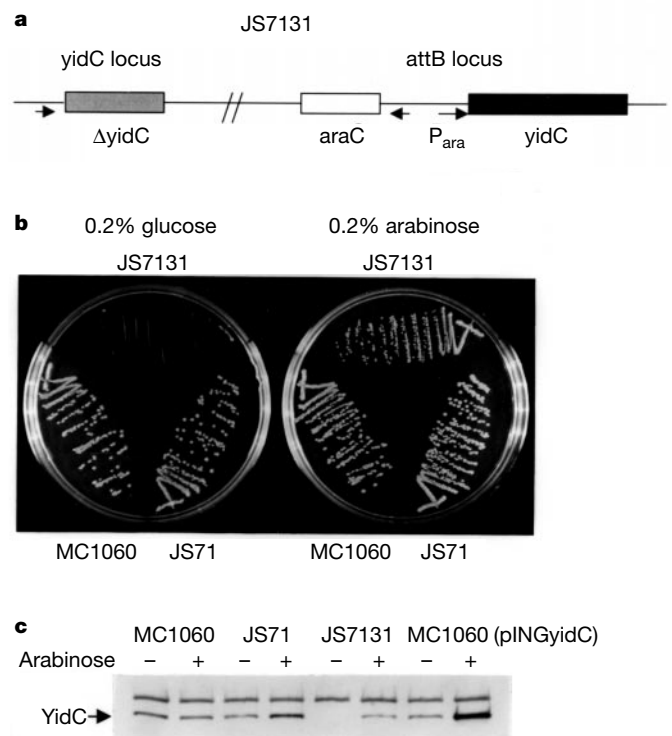


Figure 1 Genotype and phenotype of YidC-depletion strain JS7131. **a**, Schematic showing the genetic makeup of *E. coli* strain JS7131 with a *yidC* deletion and a complementary wild-type *yidC* gene under control of the *araBAD* promoter/operator. **b**, Growth of parent strains versus JS7131 on LB plates at 37 °C under conditions of repression (0.2% glucose) and induction (0.2% arabinose). **c**, Expression of YidC in JS7131 is induced with arabinose (+ lanes) and tightly repressed in the presence of 0.2% glucose (– lanes). *E. coli* strains MC1060, JS71, JS7131 or MC1060 containing the cloned *yidC* gene (pINGyidC) were grown for 3 h in LB with either 0.2% arabinose or 0.2% glucose after a 1:50 back-dilution from an overnight culture containing arabinose. The chemical amount of YidC in these strains was detected by immunoblotting with antiserum to YidC. The upper band (also precipitated by YidC antiserum) demonstrates that an equal amount of cell lysate was loaded in each lane.

membrane insertion. When YidC is present (Fig. 3a, arabinose) Procoat inserts across the membrane and is digested with proteinase K to a small carboxy-terminal fragment that is not recognized by M13 antiserum. In YidC-deficient cells (Fig. 3a, glucose), Procoat does not insert across the membrane even after a 2-min chase. To our knowledge, this shows for the first time that the membrane insertion of M13 Procoat depends on a proteinaceous factor because it requires YidC for insertion. In the same cells exhibiting a block of Procoat membrane insertion, OmpA export is essentially unaffected.

The amino-terminal region of Leader peptidase (Lep) also inserts into the inner membrane in a Sec-independent manner as demonstrated by studies using Pf3-Lep. Pf3-Lep contains the first 18 amino acids of Pf3 coat fused to the third amino acid of Lep. The introduction of an arginine at position 79 blocks Lep P2 domain translocation and allows specific assay of N-terminal translocation. As shown in Fig. 3b (arabinose), translocation of the N-terminal region of Pf3-Lep is detected by a slight shift in molecular weight in protease-treated cells. In cells grown with glucose, translocation of the N terminus of Pf3-Lep is only 60% efficient, indicating a dependence on YidC for efficient insertion (see doublet in the + proteinase K lane).

How general is this inhibitory effect on membrane protein insertion? We investigated the effect of YidC depletion on translocation of the Lep P2 domain, a Sec-dependent process. In YidC depleted cells (Fig. 3c, glucose), a significant amount of the Lep P2 domain is not inserted across the membrane after a 2-min chase. As expected, when YidC is not depleted (Fig. 3c, arabinose), the

C-terminal domain of Lep is efficiently translocated and digested with proteinase K.

Next we analysed ProW, a multi-spanning membrane protein. ProW has a long N-terminal domain that is translocated into the periplasm by a Sec-dependent pathway¹⁹ not requiring SecA for insertion¹⁸. For our studies, we analysed ProW₁₋₁₈₂-Lep, which contains the first three transmembrane segments of ProW fused to the P2 domain of Lep. We note that the membrane translocation of this fusion protein is less than 100% efficient in wild-type strain MC1061 (ref. 18). In JS7131 in the presence of arabinose to express

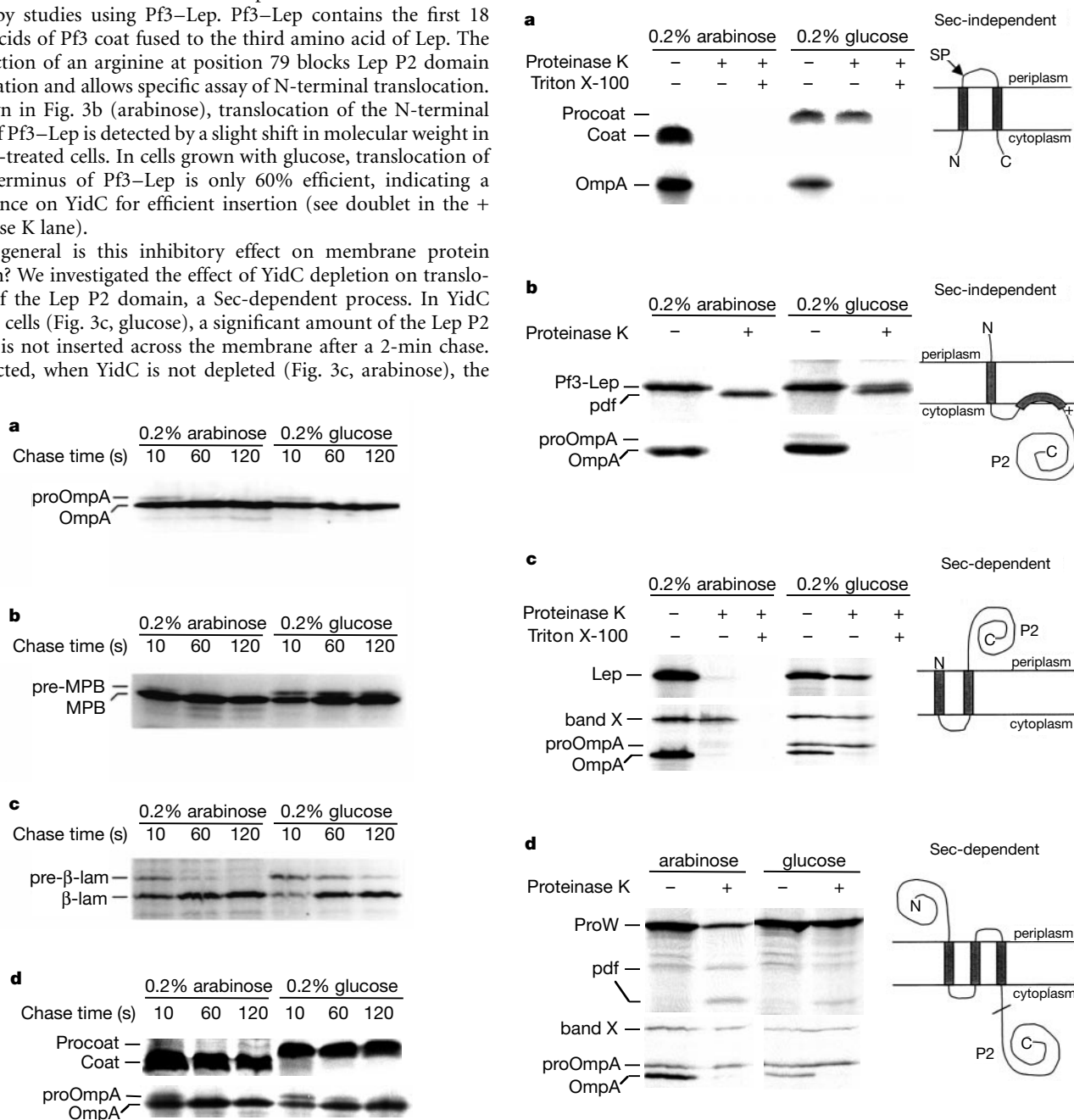


Figure 2 Effect of YidC depletion on the translocation of pre-proteins. Signal peptide processing was used as a measure of membrane translocation in JS7131 cells expressing YidC (arabinose) or depleted of YidC (glucose). **a**, Outer membrane protein A (OmpA) was expressed constitutively from the chromosome. **b**, Maltose binding protein (MBP) was induced by 1 mM IPTG (isopropyl- β -D-thiogalactopyranoside) from pMAL-p2. **c**, β -lactamase (β -lam) was expressed constitutively from pMS119. **d**, Procoat was induced by 1 mM IPTG from pMS119PC. The proteins were immunoprecipitated and analysed by SDS-PAGE and fluorography.

Figure 3 YidC is required for efficient membrane insertion of Sec-independent and Sec-dependent proteins as determined by a protease accessibility assay. **a**, Procoat. **b**, N-terminal translocation of Pf3-Lep. **c**, C-terminal translocation of the P2 domain of Lep. **d**, Translocation of the N-terminal domain of ProW. All proteins were expressed from the pMS119 vector by 1 mM IPTG addition for 10 min. After subjecting the cells to a protease accessibility assay, the proteins were immunoprecipitated and analysed by SDS-PAGE and fluorography. SP, signal peptide cleavage site; pdf, proteinase K digested fragment; band X, unidentified cytoplasmic protein precipitated by AraB antiserum which serves as a lysis control³.

YidC, 80% of the N-terminal tail of ProW₁₋₁₈₂-Lep is translocated (Fig. 3d). This is in contrast to YidC-depleted cells where only 30% of ProW₁₋₁₈₂-Lep is translocated after a 2-min chase. An additional observation is that pro-OmpA now accumulates in YidC depleted cells overexpressing Lep and ProW₁₋₁₈₂-Lep (Fig. 3c, d). We speculate that this phenomenon is due to increased traffic of a Sec-dependent membrane protein that becomes stalled in the translocase in the absence of YidC.

To obtain further information about how YidC may function, we used an *in vitro* translocation system to perform crosslinking studies where truncated messenger RNA is translated to produce truncated proteins bound to ribosomes²⁰ and a photoactivatable amino-acid

analogue, (Tmd)Phe, is introduced into the nascent protein by means of an amber suppressor tRNA²¹. These nascent chain/ribosome complexes act as intermediates in the translocation process, allowing crosslinking of the nascent chains to translocation machinery proteins. Photocrosslinking experiments using T7Lep as a model substrate demonstrate the interaction of YidC with the transmembrane segment of an inner membrane protein during insertion. T7Lep is Leader peptidase containing a T7 epitope tag at the N terminus. In the first experiment, (Tmd)Phe was incorporated at position 11 in the middle of the first transmembrane segment (H1) of T7Lep. After translation of the truncated (*M_r* 12K) nascent protein (indicated by an asterisk) in the presence of inner mem-

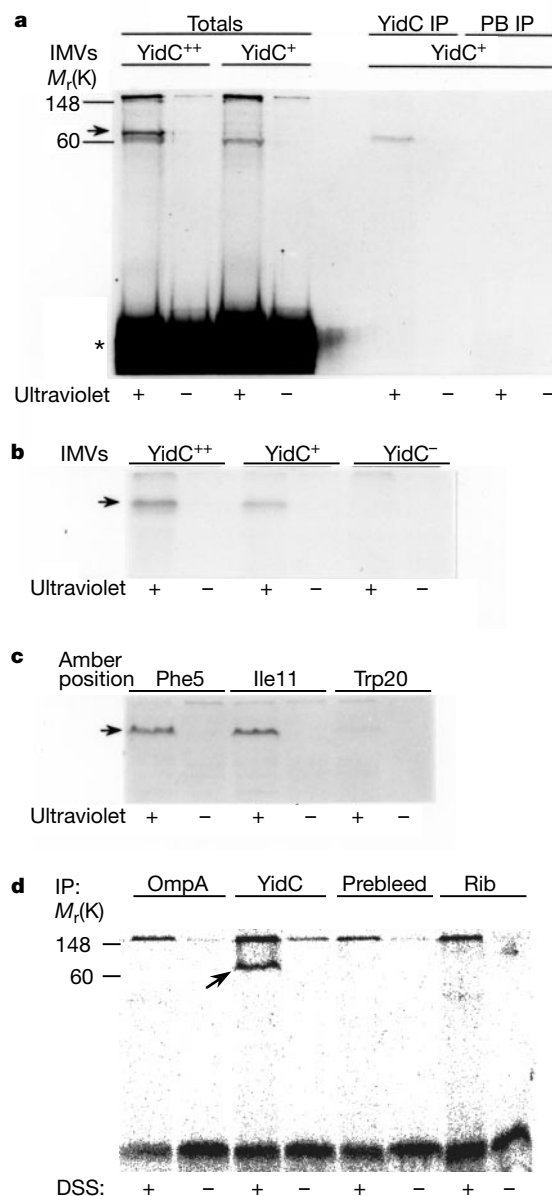


Figure 4 Crosslinking of T7Lep to YidC. The arrow indicates the photocrosslinked product, T7Lep-YidC. The molecular weight of the complex is within the range of 60–90K, which agrees with the expected molecular weight, 72K. **a**, T7Lep (111Amber) as numbered in the wild-type Lep sequence was used in a photocrosslinking reaction with IMVs prepared from YidC overproducing cells (YidC⁺⁺) or from wild-type cells (YidC⁺). The total photocrosslinking reaction products are shown on the left side. Photocrosslinking products were immunoprecipitated with YidC antiserum or prebleed serum (right side). IP, immunoprecipitation; PB, prebleed. **b**, Immunoprecipitation of the T7Lep-YidC complex from photocrosslinking reactions with IMVs containing overproduced (YidC⁺⁺), wild-type

(YidC⁺), or depleted (YidC⁻) levels of YidC. T7Lep (Ile11Amber) was used in these reactions. **c**, T7Lep mutants Phe5Amber, Ile11Amber and Trp20Amber, which allow incorporation of the photoactivatable amino-acid analogue (Tmd)Phe in different locations of the first transmembrane segment, were used in photocrosslinking reactions. After correcting for different suppression efficiencies, the degree of photocrosslinking is similar in the three locations. **d**, Chemical crosslinking of T7Lep with YidC. In this study, T7Lep without an amber mutation was used. The crosslinking reactions were immunoprecipitated with anti-YidC, anti-OmpA, anti-ribulokinase or prebleed serum.

brane vesicles (IMVs), the reactions were incubated at 4 °C with or without exposure to ultraviolet radiation. Shown in Fig. 4a (left side) are the total radiolabelled proteins with an arrow indicating the location of a crosslinked complex that is consistent with the combined molecular weights of T7Lep and YidC (72 K). Immunoprecipitation with YidC antiserum (right side) confirms the identity of the complex as T7Lep–YidC. The complex is detected at a low level with wild-type IMVs (YidC⁺), whereas complex formation is greatly increased when IMVs containing an overexpressed level of YidC are added to the translocation reaction (YidC⁺⁺).

Next we tested the amount of YidC immunoprecipitable complex when IMVs are prepared from JS7131, MC1060 and the YidC overproducing cells MC1060 (pINGyidC). As displayed in Fig. 4b, the amount of T7Lep–YidC complex is increased with IMVs prepared from the YidC overproducer, whereas no T7Lep–YidC complex is observed with YidC[−] IMVs prepared from JS7131. The fact that increased photocrosslinking to the N-terminal region of Lep is detected in YidC⁺⁺ IMVs supports the idea of YidC not being present as an integral, stoichiometric member of the Sec complex (Fig. 4a, b). Crosslinking of YidC to other positions within H1 of T7Lep was analysed by creating amber mutations at positions 5 and 20 designated as Phe5Amber and Trp20Amber, respectively (Fig. 4c). The amount of photocrosslinking to YidC was similar after correcting for different suppression efficiencies (data not shown). In addition, chemical crosslinking confirms the T7Lep–YidC interaction (Fig. 4d). These studies indicate that YidC has a direct role in membrane protein insertion.

YidC appears to cooperate with the Sec machinery for integration of proteins into the inner membrane. This is consistent with a recent study showing that YidC is associated with the purified SecYEG translocase²². A transient association with the Sec translocase would allow YidC to function in concert with other translocase subunits to insert membrane proteins into the bilayer. Possibly YidC catalyses the lateral exit of transmembrane segments from the translocase allowing integration into the lipid bilayer, a function analogous to that suggested for TRAM in the endoplasmic reticulum (ER) system^{23,24}. The crosslinking studies with Lep (Fig. 4) and FtsQ²² demonstrate that YidC interacts with hydrophobic regions of membrane proteins. As YidC may play a role in the membrane insertion of one or more of the Sec components, we cannot discount this as a factor in the effects on Sec-dependent proteins. The absolute requirement of YidC for Procoat membrane insertion suggests that YidC may function differently for proteins or domains of proteins not requiring the Sec components. A recent study of the chloroplast homologue, Albino3, provides additional evidence for a conserved Sec-independent pathway. Antibody to Albino3 was found to inhibit thylakoid membrane insertion of light-harvesting chlorophyll-binding protein, but have no effect on translocation of the Sec-dependent protein iOE33 (ref. 15). Most importantly, the absence of Sec proteins in mitochondria is in accordance with our finding of YidC being an essential and evolutionarily conserved component for the membrane insertion of Sec-independent proteins. □

Methods

Construction of the YidC-depletion strain

The *yidC* gene was amplified from chromosomal DNA of strain MC1060 which resulted in a 1,749-base-pair (bp) PCR product with an *NcoI* site at the initiation codon and an *EcoRI* site 100 bp downstream of the stop codon. The PCR product was cloned into the pCRII vector (Invitrogen) by TA cloning. After removal of the *NcoI* site at nucleotide 1,136 of the *yidC* ORF by Quikchange site-directed mutagenesis (Stratagene), the *yidC* fragment was subcloned into the *NcoI/EcoRI* sites of pRD8, thus creating pINGyidC. In this construct, *yidC* is preceded by the *lepB* upstream region to provide a Shine–Delgarno sequence, and the *araBAD* operator/promoter region including the *araC* gene to allow tightly controlled expression. The *ara–yidC* 4-kb fragment of pINGyidC was cloned into pCD13PKS²⁵ (Spec^r) in an orientation opposite to that of all other transcription within the vector. The resulting clone pCD13PKSara-yidC was permanently integrated into the *attB* site of strain MC1060 ($\Delta(codB-lac)3$, *galK16*, *galE15*, λ^- , *relA1*, *rpsL150*, *spoT1*, *hsdR2*, *ara*⁺), thus creating strain JS71 (MC1060, *attB::R6Kori*, *ParaBAD*⁺ *yidC*[−], (Spec^r)).

To generate the YidC-depletion strain JS7131, 822 bp of the *yidC* locus of strain JS71 were deleted using the knockout vector pMAKΔyidC, derived from pMAK705 (Cam^r) which is temperature-sensitive for replication²⁶. The ΔyidC mutation comprises an in-frame deletion of nucleotides 745–1566, and was generated by restriction at the naturally occurring *BstEII* and *BstBI* sites, Klenow fill-in and ligation of the blunt ends. The knockout vector also carries 370 bp of upstream DNA and the downstream *thdF* gene. pMAKΔyidC was transformed into strain JS71 and chromosomal integrants were isolated by plating on LB agar at 44 °C in the presence of 20 μg ml^{−1} chloramphenicol. Individual integrants were grown for several generations at 30 °C in LB broth with arabinose and chloramphenicol to reactivate the plasmid origin. The cultures were then diluted and plated at 30 °C to obtain individual colonies. Upon allelic exchange at the *yidC* locus, the knockout vector carried *yidC*⁺ and the strain of interest was expected to be arabinose-dependent for growth at the non-permissive temperature. This outcome was observed for ten of eighty colonies streaked at 44 °C on 0.2% arabinose-LB versus 0.2% glucose-LB plates. The final step of the strain construction was to eliminate pMAKΔyidC⁺ by a curing process combining cycloserine enrichment with growth at 44 °C. The plasmid-free strain JS7131 (JS71, ΔyidC) proved to be arabinose-dependent for growth, confirming that *yidC* is essential for cell viability.

Crosslinking

For site-specific photocrosslinking, the procedure of ref. 27 was used to probe the environment of T7Lep during translocation. Amber nonsense mutations were introduced to the T7Lep gene corresponding to positions Phe 5, Ile 11 and Trp 20 within the first transmembrane segment (H1) to allow incorporation of the photoactivatable amino-acid analogue (Tmd)Phe (1-(4'-(3-(trifluoromethyl)-3H-diazirin-3-yl)phenyl)alanine)²⁸. (Tmd)Phe-pdCPA was ligated to *E. coli* suppressor tRNA^{Asn} (ref. 29) using T4 RNA ligase (Boehringer Mannheim). Transcripts were produced by the MEGashortscript T7 kit (Ambion) from PCR fragments terminating at codon 98 of the T7Lep open reading frame. The truncated mRNA was added to an *E. coli* S100 *in vitro* translation system³⁰ to produce ³⁵S-labelled nascent proteins which remain ribosome-bound through transfer RNA. IMVs were added to the translation reactions 5 min after initiation and the reactions were allowed to proceed for 45 min at 37 °C. The IMVs were prepared from YidC overproducing cells (MC1060(pINGyidC)), wild-type cells (MC1060) and YidC deficient cells (JS7131). MC1060(pINGyidC) was grown in 0.2% arabinose-LB to overexpress YidC, and JS7131 was grown in 0.2% glucose-LB to deplete YidC. Photocrosslinking was carried out using a portable ultraviolet lamp (20 W) at 364 nm for 30 min at 4 °C. Chemical crosslinking with DSS (disuccinimidyl suberate) was carried out using the same *in vitro* system except that the truncated mRNA did not contain an amber codon. The products resulting from the crosslinking reactions were either analysed directly (totals) or immunoprecipitated before SDS–PAGE and autoradiography.

Received 10 February; accepted 19 May 2000.

1. Rapoport, T. A., Jungnickel, B. & Kutay, U. Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. *Annu. Rev. Biochem.* **65**, 271–303 (1996).
2. Phillips, G. J. & Silhavy, T. J. The *E. coli ffh* gene is necessary for viability and efficient protein export. *Nature* **359**, 744–746 (1992).
3. de Gier, J.-W. L. et al. Assembly of a cytoplasmic membrane protein in *Escherichia coli* is dependent on the signal recognition particle. *FEBS Lett.* **399**, 307–309 (1996).
4. Bernstein, H. D., Zopf, D., Freyman, D. M. & Walter, P. Functional substitution of the signal recognition particle 54kD subunit by its *Escherichia coli* homolog. *Proc. Natl Acad. Sci. USA* **90**, 5229–5233 (1993).
5. Lührbrand, N. D., Newitt, J. A. & Bernstein, H. D. The *E. coli* signal recognition particle is required for the insertion of a subset of inner membrane proteins. *Cell* **88**, 187–196 (1997).
6. Lührbrand, J. et al. An alternative protein targeting pathway in *Escherichia coli*: studies on the role of FtsY. *EMBO J.* **13**, 2289–2296 (1994).
7. Seluanov, A. & Bibi, E. FtsY, the prokaryotic signal recognition particle receptor homologue, is essential for biogenesis of membrane proteins. *J. Biol. Chem.* **272**, 2053–2055 (1997).
8. Wickner, W. & Leonard, M. R. *Escherichia coli* preprotein translocase. *J. Biol. Chem.* **22**, 29514–29516 (1996).
9. Geller, B. L. & Wickner, W. (1985). M13 procoat inserts into liposomes in the absence of other membrane proteins. *J. Biol. Chem.* **260**, 13281–13285.
10. de Gier, J.-W. L. et al. Differential use of the signal recognition particle translocase targeting pathway for inner membrane protein assembly in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **95**, 14646–14651 (1998).
11. Hell, K., Herrmann, J., Pratje, E., Neupert, W. & Stuart, R. A. Oxa1p mediates the export of the N- and C-termini of pCoxII from the mitochondrial matrix to the intermembrane space. *FEBS Lett.* **418**, 367–370 (1997).
12. Hell, K., Herrmann, J. M., Pratje, E., Neupert, W. & Stuart, R. A. Oxa1p, an essential component of the N-tail protein export machinery in mitochondria. *Proc. Natl Acad. Sci. USA* **95**, 2250–2255 (1998).
13. Sundberg, E. et al. ALBINO3, an Arabidopsis nuclear gene essential for chloroplast differentiation, encodes a chloroplast protein that shows homology to proteins present in bacterial membranes and yeast mitochondria. *Plant Cell* **9**, 717–730 (1997).
14. Bonnefoy, N., Chalvet, F., Hamel, P., Slonimski, P. P. & Dujardin, G. Oxa1, a *Saccharomyces cerevisiae* nuclear gene whose sequence is conserved from prokaryotes to eukaryotes controls cytochrome oxidase biogenesis. *J. Mol. Biol.* **239**, 201–212 (1994).
15. Moore, M., Harrison, M. S., Peterson, E. C. & Henry, R. Chloroplast Oxa1p homolog Albino3 is required for post-translational integration of the light harvesting chlorophyll-binding protein into thylakoid membranes. *J. Biol. Chem.* **275**, 1529–1532 (2000).
16. Stuart, R. A. & Neupert, W. Topogenesis of inner membrane proteins in mitochondria. *Trends Biol. Sci.* **21**, 261–267 (1996).
17. Glick, B. S. & von Heijne, G. *Saccharomyces cerevisiae* mitochondria lack a bacterial-type sec machinery. *Protein Sci.* **5**, 2651–2652 (1996).
18. Whitley, P. et al. Sec-independent translocation of a 100-residue periplasmic N-terminal tail in the *E. coli* inner membrane protein proW. *EMBO J.* **13**, 4653–4661 (1994).

19. Cristobal, S., Scotti, P., Lührink, J., von Heijne, G. & de Gier, J.-W. L. The signal recognition particle-targeting pathway does not necessarily deliver proteins to the Sec-translocase in *Escherichia coli*. *J. Biol. Chem.* **274**, 20068–20070 (1999).
20. Gilmore, R., Collins, P., Johnson, J., Kellaris, K. & Rapiejko, P. Transcription of full-length and truncated mRNA transcripts to study protein translocation across the endoplasmic reticulum. *Methods Cell Biol.* **34**, 223–237 (1991).
21. Martoglio, B., Hofmann, M. W., Brunner, J. & Dobberstein, B. The protein-conducting channel in the membrane of the endoplasmic reticulum is open laterally toward the lipid bilayer. *Cell* **18**, 207–214 (1995).
22. Scotti, P. A. et al. YidC, the *Escherichia coli* homologue of mitochondrial Oxa1p, is a component of the Sec translocase. *EMBO J.* **19**, 542–549 (2000).
23. Görlich, D., Hartmann, E., Prehn, S. & Rapoport, T. A. A protein of the endoplasmic reticulum involved early in polypeptide translocation. *Nature* **357**, 47–52 (1992).
24. Do, H., Falcone, D., Lin, J., Andrews, D. W. & Johnson, A. E. The cotranslational integration of membrane proteins into the phospholipid bilayer is a multistep process. *Cell* **85**, 369–378 (1996).
25. Platt, R., Drescher, C. D., Park, S. K. & Phillips, G. J. Genetic system for reversible integration of DNA constructs and *lacZ* gene fusions into the *Escherichia coli* chromosome. *Plasmid* **43**, 12–23 (2000).
26. Hamilton, C. M., Aldea, M., Washburn, B. K., Babitzke, P. & Kushner, S. R. New method for generating deletions and gene replacements in *Escherichia coli*. *J. Bacteriol.* **171**, 4617–4622 (1989).
27. Graf, R., Brunner, J., Dobberstein, B. & Martoglio, B. in *Cell Biology: a Laboratory Handbook* 2nd edn, Vol. 4, (ed. Celis, J. E.) 495–501 (Academic, San Diego, 1998).
28. Brunner, J. Use of photocrosslinkers in cell biology. *Trends Cell Biol.* **6**, 154–157 (1996).
29. Cload, S. T., Liu, D. R., Froland, W. A. & Schultz, P. Development of improved tRNAs for *in vitro* biosynthesis of proteins containing unnatural amino acids. *Chem. Biol.* **3**, 1033–1038 (1996).
30. Gold, L. M. & Schweiger, M. in *Methods in Enzymology* Vol. 20, (eds Moldave, K. & Grossman, L.) 537–542 (Academic, London and New York, 1971).

Acknowledgements

We thank P. G. Schultz for the *E. coli* suppressor tRNA^{Asn} gene, S. R. Kushner for knockout vector pMAK705, J.-W. de Gier for the ProW construct, J. Brunner for the photocrosslinking reagent (Tmd)Phe-pdCpA, and Matthias Muller for advice with the amber suppression studies. This work was supported by an NSF grant (to R.E.D.) and by a DFG grant (to A.K.).

Correspondence and requests for materials should be addressed to R.E.D. (e-mail: dalbey@chemistry.ohio-state.edu).

Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice

Steven E. Artandi^{*}, Sandy Chang^{*†}, Shwu-Luan Lee^{*}, Scott Alson^{*}, Geoffrey J. Gottlieb^{‡§}, Lynda Chin^{*||} & Ronald A. DePinho^{*¶}

^{*} Department of Adult Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA

[†] Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA

[‡] CPI Ameripath Laboratory, Beachwood, Ohio 44122, USA

[§] Ackerman Academy of Dermatopathology, New York, New York 10016, USA

^{||} Department of Dermatology, Harvard Medical School, Boston, Massachusetts 02115, USA

[¶] Departments of Genetics and Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA

Aged humans sustain a high rate of epithelial cancers such as carcinomas of the breast and colon, whereas mice carrying common tumour suppressor gene mutations typically develop soft tissue sarcomas and lymphomas. Among the many factors that may contribute to this species variance are differences in telomere length and regulation. Telomeres comprise the nucleoprotein complexes that cap the ends of eukaryotic chromosomes and are maintained by the reverse transcriptase, telomerase¹. In human cells, insufficient levels of telomerase lead to telomere attrition with cell division in culture² and possibly with ageing and tumorigenesis *in vivo*^{3–5}. In contrast, critical reduction in telomere length is not observed in the mouse owing to promiscuous telomerase expression and long telomeres^{6–10}. Here we

provide evidence that telomere attrition in ageing telomerase-deficient p53 mutant mice promotes the development of epithelial cancers by a process of fusion-bridge breakage that leads to the formation of complex non-reciprocal translocations—a classical cytogenetic feature of human carcinomas. Our data suggest a model in which telomere dysfunction brought about by continual epithelial renewal during life generates the massive ploidy changes associated with the development of epithelial cancers.

Chromosomal rearrangement mechanisms are intimately linked to cancer development and are thought to generate the numerous gains and losses of segments of chromosomes needed for epithelial carcinogenesis. These wholesale and complex rearrangements typically occur early, by the carcinoma-*in-situ* stage, but appear to be largely unchanged with progression to invasive and metastatic disease^{11–13}. Additional genomic alterations and mutations no doubt accrue during tumour progression, but these data suggest that the instability of cancer genomes is episodic, primarily occurring early during carcinogenesis and resolving to relative stability in advanced malignancy. These genomic changes are evident by the stage when telomerase is first activated^{14,15}, which suggests that an early and transient period of telomere dysfunction could contribute to complex genomic alterations encountered in mature epithelial cancers.

Mice lacking the RNA component of telomerase (mTERC) exhibit progressive telomere shortening and ultimately chromosomal instability (end-to-end fusions) as a function of age and of successive generational matings^{16,17}. Paradoxically, although telomerase facilitates oncogenic transformation of cultured human cells¹⁸, telomere shortening in ageing mTERC^{−/−} mice is associated with increased rates of cancer, suggesting that the genetic instability

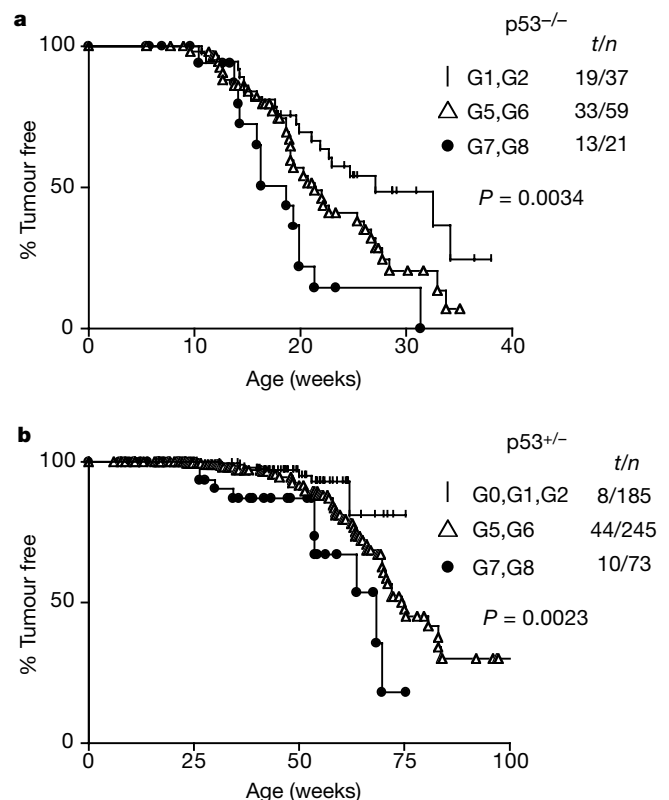


Figure 1 Kaplan-Meier analysis of tumour incidence in p53 mutant mice divided on the basis of generation of telomerase deficiency. **a**, p53^{−/−} mice. The number of tumours identified (t) and the total number of mice (n) in each cohort is indicated. Hatched line, G1–G2 mTERC^{−/−}; triangles, G5–G6 mTERC^{−/−}; circles, G7–G8 mTERC^{−/−}. **b**, p53^{+/-} mice. Hatched line, mTERC^{+/+}, mTERC^{+/-} or G1–G2 mTERC^{−/−}; triangles, G5–G6 mTERC^{−/−}; circles, G7–G8 mTERC^{−/−}.

Table 1 Histological diagnoses of spontaneous tumours

Histological type	Telomere function intact		Telomere dysfunction	
	No.	% of tumours	No.	% of tumours
Tumour spectrum in p53 ^{-/-} mice				
Thymic lymphoma	23	64	24	54
Other lymphoma	3	8	2	4
Sarcoma	9	25	15	33
Glioma, spinal cord	1	3	0	0
Adenocarcinoma	0	0	4	9
Total	36		100	45
Tumour spectrum in p53 ^{+/+} mice*				
Breast adenocarcinoma	0	0	10	10
Squamous cell carcinoma	0	0	21	22
Gastrointestinal adenocarcinoma	0	0	19	20
Other carcinoma	0	0	3	3
Osteosarcoma	2	25	17	18
Angiosarcoma	0	0	4	4
Other sarcoma	4	50	11	11
Ovarian stromal cell tumour	0	0	4	4
Lymphoma	2	25	8	8
Total	8	100	97	100

* p53^{-/-} mice were compared between two cohorts: one with intact telomeres (mTERC^{+/+}, mTERC^{-/-}, G1–G2 mTERC^{-/-}) and one with severe telomere dysfunction (G5–G8 mTERC^{-/-}).

associated with telomere dysfunction can facilitate transformation *in vivo*¹⁹. Moreover, loss of p53 improves the survival of cells with telomere dysfunction, and thus has a permissive role in the generation of end-to-end chromosomal fusions and aneuploidy. In the setting of p53 deficiency, telomere dysfunction confers a greater susceptibility to transformation by cellular oncogenes²⁰. Despite these observations, the precise mechanism by which telomere

dysfunction facilitates cancer and the implications of these tissue culture experiments for tumorigenesis *in vivo* remained unclear.

Although telomerase is reactivated in most mature human cancers, we sought to model the early stages of tumorigenesis in which telomere shortening and dysfunction might impair chromosomal integrity. We monitored the cancer phenotype and cytogenetics of large cohorts of telomerase-deficient p53 mutant mice. In generation one (G1) and generation two (G2) mTERC^{-/-} p53^{-/-} mice that lack telomerase but retain normal telomere function, median tumour incidence was similar to that of mTERC^{+/+} p53^{-/-} and mTERC^{+/+} p53^{-/-} controls (27.1 compared with 24.9 weeks, $P = 0.212$; data not shown). In contrast, in later mTERC^{-/-} generations the progressive decline in telomere function correlated with decreasing tumour latency: 27.1 weeks for G1–G2, 21.3 weeks for G5–G6, and 18.6 weeks for G7–G8 mTERC^{-/-} p53^{-/-} cohorts ($P = 0.0034$; Fig. 1a). Similarly, in p53^{+/+} cohorts median tumour latency decreased progressively as telomeres became short and dysfunctional ($P = 0.0023$; Fig. 1b). Seven out of nine tumours of late generation mTERC^{-/-} p53^{+/+} mice exhibited loss of the remaining p53 wild-type allele (data not shown). Thus, telomere dysfunction, rather than loss of telomerase *per se*, cooperates with p53 deficiency to accelerate tumorigenesis *in vivo*, in agreement with previous transformation studies in cell culture²⁰.

Sarcomas and lymphomas dominated the tumour spectrum of late generation mTERC^{-/-} p53^{-/-} mice (Table 1), similar to previous reports of p53^{-/-} mice^{21,22}. However, the emergence of adenocarcinomas in late generation mTERC^{-/-} p53^{-/-} mice (9%), compared with none in early generation controls, suggested that telomere dysfunction may be involved in promoting epithelial carcinogenesis. Because p53^{-/-} mice succumb rapidly to lymphoid and mesenchymal cancers, we reasoned that the much longer tumour latency of p53^{+/+} mice^{21,22} might reveal an impact of age-dependent

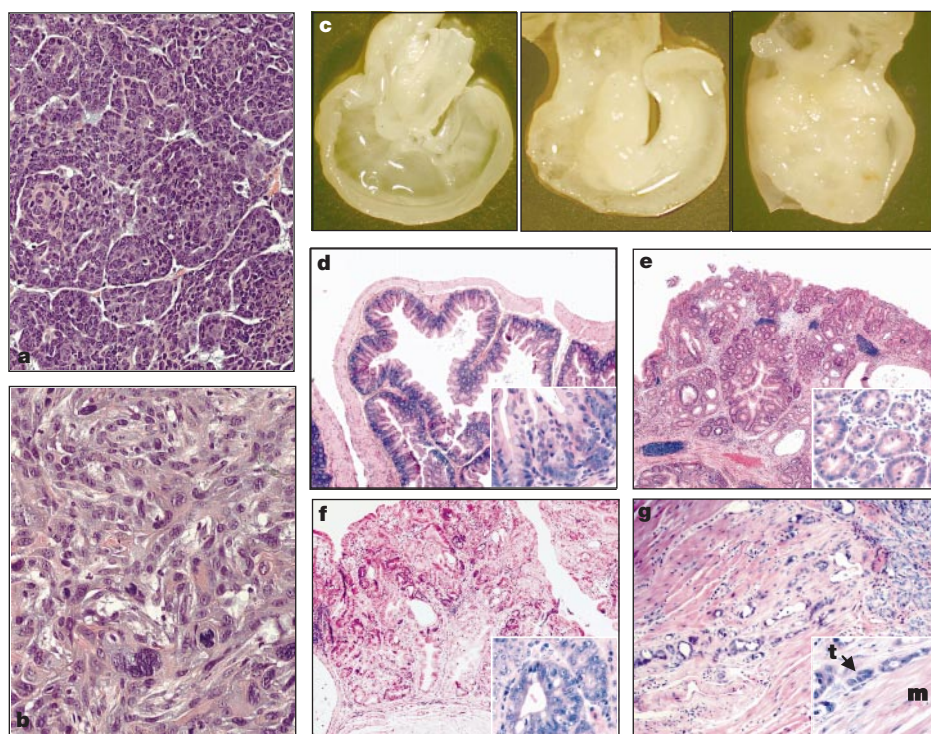


Figure 2 Histology of epithelial cancers in mice with telomere dysfunction. **a**, Breast cancer, G5 mTERC^{-/-} p53^{+/+} mouse; **b**, squamous cell carcinoma, G6 mTERC^{-/-} p53^{+/+} mouse. H&E stain, original magnification $\times 20$. **c**, Gross view of caeca from mTERC^{-/-} p53^{+/+} (left), G6 mTERC^{-/-} p53^{+/+} (middle), and G5 mTERC^{-/-} p53^{+/+} (right). **d**, Histology of normal caecum, mTERC^{-/-} p53^{+/+}, shows typical colonic villi and ordered nuclei. H&E stain, original magnification $\times 4$; inset, $\times 20$. **e**, Adenomatous polyp in the caecum, G5

mTERC^{-/-} p53^{+/+}. Inset, glands remain round with basal nuclei. H&E stain, original magnification $\times 4$; inset, $\times 20$. **f**, Caecal adenocarcinoma, G5 mTERC^{-/-} p53^{+/+}. Inset, disordered glands and pleomorphic nuclei. Original magnification $\times 4$, inset $\times 20$. **g**, Invasive adenocarcinoma of colon, G5 mTERC^{-/-} p53^{+/+}. Inset, tumour cells (t) with poor glandular organization invading muscle fibres (m). H&E stain, original magnification $\times 10$; inset, $\times 20$.

telomere attrition on tumorigenesis in self-renewing epithelial compartments. We subjected 97 tumours arising in 81 late generation mTERC^{-/-} p53^{+/-} mice to detailed histological and tumour marker analysis (Table 1). Carcinomas represented the largest class of clinically apparent tumours (55%, but occult carcinomas occurred in all mice, see below), exceeding sarcomas (37%) and lymphomas (8%).

Adenocarcinoma of the breast presented as solitary subcutaneous masses along the milk line in female mice only, and, on the histological level, showed carcinoma cells growing in nested patterns (Fig. 2a). Squamous cell carcinomas of the skin showed characteristic morphology (keratin pearls and intercellular bridges, Fig. 2b), and immunoreactivity to cytokeratins—markers commonly used to diagnose human epithelial cancers (see Supplementary Information). A subset of mice ($n = 16$) exhibited either distension of the gastrointestinal tract or caecal enlargement. In roughly 75% of these mice, gross examination of the caecum revealed a macroscopic polyp, usually extending along the luminal wall of the caecum without a discernible stalk (sessile polyp), findings typical of caecal polyps in humans (Fig. 2c). Histologically, these lesions showed a spectrum of abnormalities from adenomatous proliferation to frank adenocarcinoma, including abnormal glandular architecture, pleomorphic nuclei and frequent mitotic figures (Fig. 2d–f). In addition, adenocarcinomas of the small bowel ($n = 2$) and colon ($n = 2$) were detected in this group of mice and all four tumours exhibited clear invasion of muscle (Fig. 2g). To determine the frequency of gastrointestinal pathology in the population, 10 mice from the G5–G7 mTERC^{-/-} p53^{+/-} cohort were killed and the caeca analysed (mean age 63 weeks). On gross examination, 6 out of 10 mice had either macroscopic polyps or thickened and irregular mucosa, and all 10 showed clear histological abnormalities including adenomatous hyperplasia and adenocarcinoma. Eleven p53^{+/-} mice with intact telomeres (mTERC^{+/+}, mTERC^{+/+} and G1–G2 mTERC^{-/-}) were analysed as controls

(mean age 50 weeks). The caeca from all 11 controls were normal on gross and histological examination. Despite the large size and invasiveness of these carcinomas, metastases were not detected.

To determine the mechanism by which telomere dysfunction accelerates tumour onset and shifts tumour spectrum, we analysed telomere maintenance and chromosomal structure in spontaneous cancers from mTERC^{-/-} p53 mutant mice. Telomere PNA-fluorescent *in situ* hybridization (FISH) on metaphase spreads of lymphomas and breast cancers derived from late generation mTERC^{-/-} mice revealed a high frequency of chromosome ends lacking telomere signal (signal-free ends) and numerous end-to-end fusions. In contrast, signal-free ends and fusions were uncommon in lymphomas from G1 mTERC^{-/-} p53^{-/-} mice (see Supplementary Information). Microscopic analysis revealed frequent anaphase bridges and lagging chromosomes in late generation mTERC^{-/-} p53^{+/-} breast carcinomas, but none in tumours arising in early generation mTERC^{-/-} mice (see Supplementary Information). These bridges represent dicentric chromosomes pulled to opposite poles by the kinetochore/spindle apparatus leading to chromosomal breakage^{23,24}. Perhaps as evidence of increased breakage, small chromosomal fragments were detected in metaphase spreads of late generation mTERC^{-/-} p53 mutant lymphomas and breast tumours, but not in G1 mTERC^{-/-} p53 mutant lymphomas (Fig. 3d, arrows). Together, these data provide direct evidence for persistent telomere dysfunction in a subpopulation of tumour cells.

The availability of p53 null lymphomas from mice with normal telomeres and from mice with severe telomere dysfunction allowed us to study the impact of telomere dysfunction on chromosomal structure in the same tumour type. Spectral karyotyping (SKY)²⁵ revealed that p53-null lymphomas with intact telomere function, including one mTERC^{+/+} and one G1 mTERC^{-/-} tumours, exhibited mild aneuploidy, but no fusions or translocations (Fig. 3a). These findings are consistent with studies^{26,27} indicating that the mechanism of p53-deficient lymphomagenesis does not involve translocation to

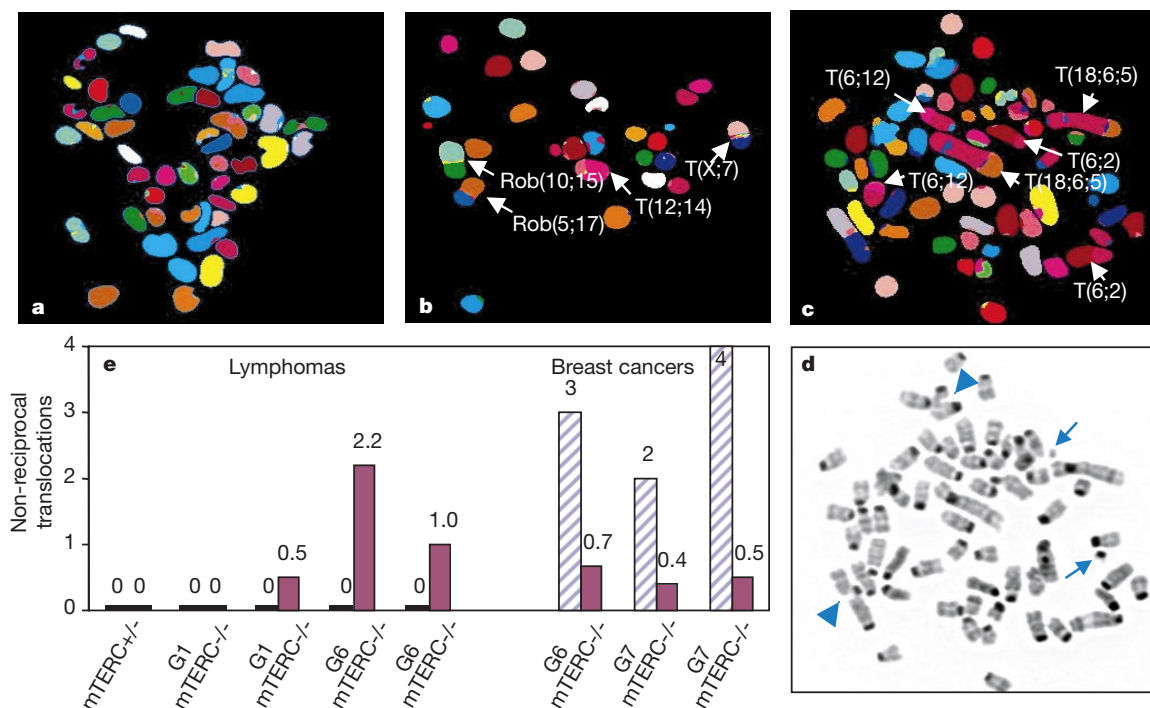


Figure 3 Spectral karyotyping. **a**, Metaphase spread, G1 mTERC^{-/-} p53^{+/-} lymphoma, showing no chromosomal rearrangements. **b**, G6 mTERC^{-/-} p53^{+/-} lymphoma, showing the presence of NRTs (T) and end-to-end chromosomal fusions (Rob). **c**, G7 mTERC^{-/-} p53^{+/-} breast carcinoma, showed aneuploidy and several complex NRTs, some of which involved multiple chromosomes. **d**, Reverse DAPI image of the same metaphase as in **c**

reveals marked genomic instability including chromatid breaks (arrowheads) and chromosome fragments (arrows). **e**, For each lymphoma and breast cancer analysed, the number of clonal NRTs (blue, striped) and non-clonal NRTs (pink) per metaphase is represented in a bar graph.

immunoglobulin gene loci. A second G1 mTERC^{-/-} p53^{-/-} lymphoma, originating 10 weeks later than the first, exhibited low numbers of end-to-end fusions—evidence for mild telomere dysfunction—and showed the emergence of interstitial chromosomal translocations of the non-reciprocal type. Notably, two G6 mTERC^{-/-} p53^{-/-} lymphomas exhibited numerous end-to-end fusions, indicating severe telomere dysfunction, and frequent non-reciprocal translocations (NRTs) (Fig. 3b, e). These data indicate that loss of telomere integrity due to progressive telomere shortening facilitates formation of NRTs.

In the lymphomas, both the end-to-end fusions and NRTs were non-recurrent; that is, consistent changes were not seen among multiple metaphases from the same tumour. SKY analysis of spontaneous breast tumours in G6 and G7 mTERC^{-/-} p53^{-/-} mice also revealed numerous complex NRTs and marked aneuploidy in all cases (Fig. 3c, d). In contrast to the lymphomas, each breast cancer exhibited recurrent NRTs, indicating a clonal origin of these rearrangements (Fig. 3e). In addition, non-recurrent NRTs were present at lower frequency in the breast cancer metaphases. These cytogenetics are remarkably similar to those of human breast cancers, which typically exhibit both clonal and non-clonal NRTs²⁸. Because breast cancers occurred in late generation mTERC^{-/-} p53^{-/-} mice only, tumours from transgenic models of mammary carcinoma were the best available controls to study chromosome structure in murine breast cancers with intact telomeres. Preliminary analysis of breast cancers from MMTV-Wnt-1 transgenic mice²⁹ revealed modest levels of aneuploidy and no NRTs (S.E.A. *et al.*, unpublished data). These results suggest that NRTs are not required for carcinogenesis in mouse mammary epithelial cells driven by a strong oncogenic stimulus.

Here we have shown that telomere dysfunction in p53 mutant mice promotes generation of NRTs, accelerates carcinogenesis and shifts the tumour spectrum toward one dominated by carcinomas. Non-reciprocal translocations have oncogenic potential in two ways: first, by carrying chimaeric or deregulated oncogenes at their breakpoints, as do reciprocal translocations; and second, by widely altering gene dosage. This latter process—the gain or loss of genetic information—makes NRTs fundamentally different from their balanced reciprocal counterparts. It is striking that human carcinoma cytogenetics are characterized by NRTs and that these rearrangements are often clonal within a tumour in similar fashion to the breast cancers that we describe here. In contrast to these adult tumours, NRTs are much less common in paediatric cancers such as lymphomas, leukemias and sarcomas, tumours in which reciprocal translocations frequently define the disease in molecular terms^{30,31}.

Propagation of a dicentric chromosome in plants has been shown to lead to structural gains and losses at the terminus of sister chromatids by repeated rounds of fusion-bridge breakage³². We propose that compromise of telomere integrity can lead to NRTs by a similar model of fusion-breakage translocation. Formation of dicentric chromosomes caused by telomere dysfunction may lead to chromosome breakage, generating recombinogenic free DNA ends that invade other chromosomes, yielding NRTs. Finally, the findings that telomere dysfunction generates these non-reciprocal changes and promotes carcinoma development provide a framework for understanding the well recognized but poorly understood paradox of episodic instability in human carcinomas. Telomere shortening in self-renewing compartments during life may compromise telomere function, leading to widespread gains and losses of chromosomal regions, and thus facilitating selection for an optimal genomic profile during tumour initiation. Further maturation is probably enhanced by quelling the extreme genomic instability of telomere dysfunction through telomerase reactivation. Our data suggest that 'telomeric instability' may initiate the neoplastic process and that telomerase reactivation may pave the transition to a more stable genome in which more subtle changes promote tumour progression. □

Methods

Mating scheme for generating mice

mTERC^{-/-} mice were mated with p53^{-/-} mice (Jackson Labs) to generate double heterozygotes (mTERC^{+/-} p53^{+/-}). Both strains are of mixed genetic background, primarily 129SV and C57Bl/6. These mice were intercrossed to generate G1 mTERC^{-/-} p53^{-/-} mice; G1 mTERC^{-/-} p53^{-/-} intercrosses yielded G2 mTERC^{-/-} p53^{-/-} mice and so on until G8. We used randomized cousin mating schemes to maintain genetic heterogeneity and prevent the generation of substrains³⁷.

Tumour incidence and analysis

Mice were examined closely for evidence of ill health or overt tumour growth. Mice were killed if profoundly ill or if external tumours exceeded 2 cm in diameter and scored as a death in survival analysis. Only those animals with histologically proven cancer were scored as an event in the tumour incidence Kaplan–Meier analysis. Statistical significance was measured using the log-rank test. All mice were subjected to extensive autopsy and gross examination. Tumours were fixed in 10% formalin and embedded in paraffin. Five-µm paraffin sections were stained with haematoxylin and eosin.

Spectral karyotyping

Spectral karyotyping of lymphoma and breast cancer cell lines was done essentially as described²⁵. At least 12 metaphases were analysed for each of the breast carcinomas, and between 5 and 10 metaphases were analysed for each of the lymphomas. Structural aberrations were determined to be clonal if found in two or more metaphases.

Received 17 April; accepted 24 May 2000.

- Sedivy, J. M. Can ends justify the means?: telomeres and the mechanisms of replicative senescence and immortalization in mammalian cells. *Proc. Natl Acad. Sci. USA* **95**, 9078–9081 (1998).
- Harley, C. B., Futcher, A. B. & Greider, C. W. Telomeres shorten during ageing of human fibroblasts. *Nature* **345**, 458–460 (1990).
- Hastie, N. D. *et al.* Telomere reduction in human colorectal carcinoma and with ageing. *Nature* **346**, 866–868 (1990).
- Allsopp, R. C. *et al.* Telomere length predicts replicative capacity of human fibroblasts. *Proc. Natl Acad. Sci. USA* **89**, 10114–10118 (1992).
- Chadeneau, C., Hay, K., Hirte, H. W., Gallinger, S. & Bacchetti, S. Telomerase activity associated with acquisition of malignancy in human colorectal cancer. *Cancer Res.* **55**, 2533–2536 (1995).
- Kipling, D. & Cooke, H. J. Hypervariable ultra-long telomeres in mice. *Nature* **347**, 400–402 (1990).
- Prowse, K. R. & Greider, C. W. Developmental and tissue-specific regulation of mouse telomerase and telomere length. *Proc. Natl Acad. Sci. USA* **92**, 4818–4822 (1995).
- Broccoli, D., Godley, L. A., Donehower, L. A., Varmus, H. E. & de Lange, T. Telomerase activation in mouse mammary tumours: lack of detectable telomere shortening and evidence for regulation of telomerase RNA with cell proliferation. *Mol. Cell. Biol.* **16**, 3765–3772 (1996).
- Greenberg, R. A., Allsopp, R. C., Chin, L., Morin, G. B. & DePinho, R. A. Expression of mouse telomerase reverse transcriptase during development, differentiation and proliferation. *Oncogene* **16**, 1723–1730 (1998).
- Martin-Rivera, L., Herrera, E., Albar, J. & Blasco, M. A. Expression of mouse telomerase catalytic subunit in embryos and adult tissues. *Proc. Natl Acad. Sci. USA* **95**, 10471–10476 (1998).
- Nishizaki, T. *et al.* Genetic alterations in primary breast cancers and their metastases: direct comparison using modified comparative genomic hybridization. *Genes Chromosomes Cancer* **19**, 267–272 (1997).
- Buerger, H. *et al.* Comparative genomic hybridization of ductal carcinoma in situ of the breast: evidence of multiple genetic pathways. *J. Pathol.* **187**, 396–402 (1999).
- Al-Mulla, F., Keith, W. N., Pickford, I. R., Going, J. J. & Birnie, G. D. Comparative genomic hybridization analysis of primary colorectal carcinomas and their synchronous metastases. *Genes Chromosomes Cancer* **24**, 306–314 (1999).
- Kolquist, K. A. *et al.* Expression of TERT in early premalignant lesions and a subset of cells in normal tissues. *Nature Genet.* **19**, 182–186 (1998).
- Shpitz, B. *et al.* Telomerase activity in ductal carcinoma in situ of the breast. *Breast Cancer Res. Treat.* **58**, 65–69 (1999).
- Blasco, M. A. *et al.* Telomere shortening and tumour formation by mouse cells lacking telomerase RNA. *Cell* **91**, 25–34 (1997).
- Lee, H. W. *et al.* Essential role of mouse telomerase in highly proliferative organs. *Nature* **392**, 569–574 (1998).
- Hahn, W. C. *et al.* Creation of human tumour cells with defined genetic elements. *Nature* **400**, 464–468 (1999).
- Rudolph, K. L. *et al.* Longevity, stress response, and cancer in aging telomerase-deficient mice. *Cell* **96**, 701–712 (1999).
- Chin, L. *et al.* p53 deficiency rescues the adverse effects of telomere loss and cooperates with telomere dysfunction to accelerate carcinogenesis. *Cell* **97**, 527–538 (1999).
- Harvey, M. *et al.* Spontaneous and carcinogen-induced tumorigenesis in p53-deficient mice. *Nature Genet.* **5**, 225–229 (1993).
- Jacks, T. *et al.* Tumour spectrum analysis in p53-mutant mice. *Curr. Biol.* **4**, 1–7 (1994).
- Kirk, K. E., Harmon, B. P., Reichardt, I. K., Sedat, J. W. & Blackburn, E. H. Block in anaphase chromosome separation caused by a telomerase template mutation. *Science* **275**, 1478–1481 (1997).
- van Steensel, B., Smogorzewska, A. & de Lange, T. TRF2 protects human telomeres from end-to-end fusions. *Cell* **92**, 401–413 (1998).
- Liyanage, M. *et al.* Multicolour spectral karyotyping of mouse chromosomes. *Nature Genet.* **14**, 312–315 (1996).
- Nacht, M. & Jacks, T. V(D)J recombination is not required for the development of lymphoma in p53-deficient mice. *Cell Growth Differ.* **9**, 131–138 (1998).
- Liao, M. J. *et al.* No requirement for V(D)J recombination in p53-deficient thymic lymphoma. *Mol. Cell. Biol.* **18**, 3495–3501 (1998).
- Heim, S. & Mitelman, F. in *Cancer Cytogenetics* 369–388 (Wiley-Liss, New York, 1995).

29. Tsukamoto, A. S., Grosschedl, R., Guzman, R. C., Parslow, T. & Varmus, H. E. Expression of the int-1 gene in transgenic mice is associated with mammary gland hyperplasia and adenocarcinomas in male and female mice. *Cell* **55**, 619–625 (1988).
30. Atkin, N. B. Lack of reciprocal translocations in carcinomas. *Cancer Genet. Cytogenet.* **21**, 275–278 (1986).
31. Hilgenfeld, E., Padilla-Nash, H., Schrock, E. & Ried, T. Analysis of B-cell neoplasias by spectral karyotyping (SKY). *Curr. Top. Microbiol. Immunol.* **246**, 169–174 (1999).
32. McClintock, B. The stability of broken ends of chromosomes in *Zea mays*. *Genetics* **26**, 234–282 (1941).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank D. Castrillon for advice on tumour analysis; S. Ye for technical assistance; M. Novotny and C. Lam for technical assistance with histopathology; L. Rudolph, K. Wong, N. Sharpless, C. Khoo and N. Schreiber-Agus for critical reading of the manuscript. This work was supported by grants from the NIH to R.A.D. S.E.A. is a Howard Hughes Physician Postdoctoral Fellow and supported by a grant from the Massachusetts Department of Public Health. L.C. is a V foundation scholar and is supported by a NIH Mentored Clinician Scientist Award. R.A.D. is an American Cancer Society Professor. SKY analysis was performed in the Arthur and Rochelle Belfer Cancer Genomic Center at the Dana-Farber Cancer Institute.

Correspondence and requests for materials should be addressed to R.A.D. (e-mail: ron_depinho@dfci.harvard.edu).

Helix deformation is coupled to vectorial proton transport in the photocycle of bacteriorhodopsin

Antoine Royant^{†‡§}, Karl Edman[§], Thomas Ursby[†],
Eva Pebay-Peyroula^{*}, Ehud M. Landau[¶] & Richard Neutze[§]

^{*} Institut de Biologie Structurale, CEA-CNRS-Université Joseph Fourier, UMR 5075, 41 rue Jules Horowitz, F-38027 Grenoble Cedex 1, France

[†] European Synchrotron Radiation Facility, 6 rue Jules Horowitz, BP 220, F-38043 Grenoble Cedex, France

[§] Department of Biochemistry, Uppsala University, Biomedical Centre, Box 576, S-751 23 Uppsala, Sweden

[¶] Biozentrum, University of Basel, Klingelbergstrasse 70, 4056 Basel, Switzerland

[‡] These authors contributed equally to this work.

A wide variety of mechanisms are used to generate a proton-motive potential across cell membranes, a function lying at the heart of bioenergetics. Bacteriorhodopsin, the simplest known proton pump¹, provides a paradigm for understanding this process. Here we report, at 2.1 Å resolution, the structural changes in bacteriorhodopsin immediately preceding the primary proton transfer event in its photocycle. The early structural rearrangements² propagate from the protein's core towards the extracellular surface, disrupting the network of hydrogen-bonded water molecules that stabilizes helix C in the ground state. Concomitantly, a bend of this helix enables the negatively charged³ primary proton acceptor, Asp85, to approach closer to the positively charged primary proton donor, the Schiff base. The primary proton transfer event would then neutralize these two groups, cancelling their electrostatic attraction and facilitating a relaxation of helix C to a less strained geometry. Reprotonation of the Schiff base by Asp85 would thereby be impeded, ensuring vectorial proton transport. Structural rearrangements also occur near the protein's surface, aiding proton release to the extracellular medium.

Photoisomerisation of the bacteriorhodopsin (bR) chromophore from an all-*trans* retinal, covalently linked to Lys 216 through

a protonated Schiff base, to the 13-*cis* configuration is the initial event in the proton pumping mechanism. During its photocycle bacteriorhodopsin passes through a series of well characterized spectral intermediates, the simplest representation of which is $bR_{570} \rightarrow K_{590} \leftrightarrow L_{550} \leftrightarrow M_{412} \leftrightarrow N_{560} \leftrightarrow O_{640} \rightarrow bR_{570}$. The initial proton transfer occurs in the $L_{550} \rightarrow M_{412}$ transition. To elucidate the structural changes facilitating this transfer, we established conditions under which a high population of the low-temperature L intermediate (L_{LT}) builds up within crystals of wild-type bacteriorhodopsin grown in a lipidic cubic phase⁴, and determined the associated structural rearrangements.

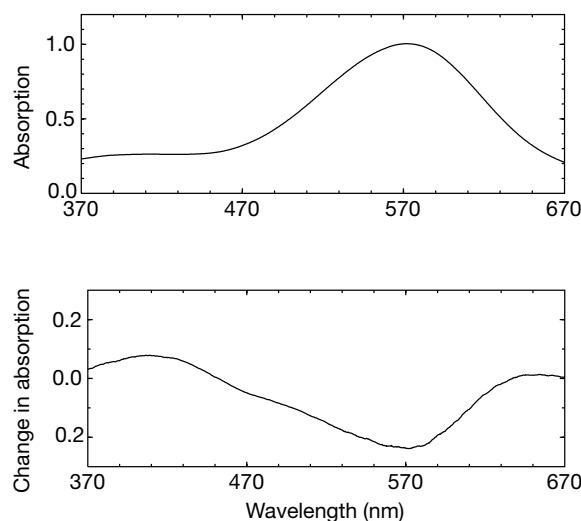


Figure 1 Spectral characterization of a bacteriorhodopsin (bR) intermediate trapped within a single crystal. **a**, Absorption spectrum of a typical light-adapted bR crystal in the ground state at 170 K. **b**, Difference spectrum between the spectrum of the crystal obtained 40 s after being illuminated for 30 s with green light ($\lambda = 532$ nm), and the spectrum shown in **a**. This spectrum is characteristic of a build up of the low-temperature L state (L_{LT}), with a small contribution from a deprotonated state.

Table 1 Diffraction data and refinement statistics

Data set	Ground*	Excited
Resolution (Å)	38–1.95	30–2.1
No. of observations	86,583	73,411
No. of unique reflections	16,826	13,088
Completeness (%) (outer shell)†	99.4 (99.1)	96.6 (99.2)
R_{sym} (%)‡ (outer shell)	4.1 (25.2)	6.3 (52.1)
mean $I/\sigma(I)$ (outer shell)	12.5 (2.9)	9.0 (1.4)
Unit cell (Å)	60.84, 60.84, 110.49	60.96, 60.96, 109.97
Twinning (%)	< 2	26 ± 2
Conformers		Ground/ L_{LT}
Occupancy (%)		30/70
Number of atoms¶		
Total		1,798/1,776
Protein		1,752/1,752
Retinal		20/20
Water		26/4
R_{cryst} (%)# (outer shell)**		26.48 (29.35)
R_{free} (%)‡‡ (outer shell)		28.82 (30.52)
r.m.s. deviation of bond lengths (Å)		0.0075
r.m.s. deviation of bond angles (°)		1.15

* This corresponds to a new integration of the data that yielded the observations from which PDB entry 1qhi was refined¹⁰.

† Outer shells were 2.06/1.95 and 2.21/2.10 Å for the ground state and the excited state, respectively. R_{merge} on the intensities between the ground state and the excited state corrected for twinning was 26.2%.

‡ $R_{\text{sym}} = \sum_i \sum_h |I_{ij} - \langle I_{ij} \rangle| / \sum_i \sum_h I_{ij}$

§ The space group is $P6_3$.

|| The conformer corresponding to the Ground state was strictly fixed during refinement.

¶ Numbers correspond to the non-refined Ground state conformer and to the refined L_{LT} state conformer, respectively.

$R_{\text{cryst}} = \sum_i |F_{\text{obs}}(h)| - |F_{\text{calc}}(h)| / \sum_i |F_{\text{obs}}(h)|$

** Refinement outer shell was 2.17/2.10 Å.

‡‡ Calculated from a set of 5% randomly selected reflections that were excluded from refinement. This set of reflections is identical to the one used in the ground state refinement. The R_{free} for the ground state refinement was 24.5%.

§ Present address: Department of Physiology and Biophysics and Sealy Center for Structural Biology, University of Texas Medical Branch, 301 University Boulevard, Galveston, Texas 77555-0641, USA.

Careful control of the temperature of a crystal during illumination has enabled specific structural intermediates of bacteriorhodopsin² and other proteins^{5,6} to be determined. An intermediate trapping protocol for L_{LT} similar to those used when trapping the L state in purple membrane^{7,8}, was established using single crystal microspectrophotometry⁹. Figure 1a shows the ground state spectrum of a bacteriorhodopsin crystal maintained at 170 K by a nitrogen gas stream. After illumination for 30 s with green light followed by 40 s without illumination, we observed the difference spectrum shown in Fig. 1b. This shows a build up of the L_{LT}

intermediate, which also contains a small contribution from a deprotonated state. The delay without illumination before flash-freezing ensured that a positive difference spectrum feature at 630 nm, indicating a contribution from the low-temperature K intermediate² (K_{LT}), decayed entirely. Crystals were subsequently flash-frozen in liquid nitrogen and monochromatic X-ray diffraction data sets (F_{exc}) were recorded (Table 1). The main conclusions described below were drawn from ($F_{\text{exc}} - F_{\text{ground}}$) difference Fourier maps and were confirmed by refinement (Table 1).

Figure 2 shows a global view (in stereo) of the observed electron

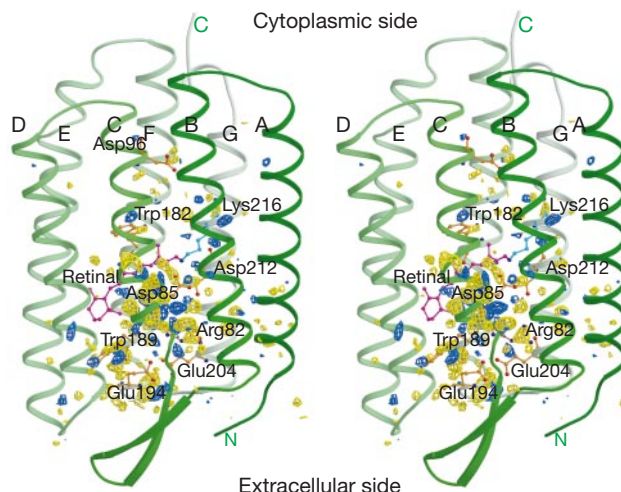


Figure 2 Structural changes in bacteriorhodopsin resulting from photoexcitation at 170 K. Overview in stereo of the $F_{\text{exc}} - F_{\text{ground}}$ difference electron-density map contoured to 3.4σ and overlaid on the ground-state bR model. Yellow, negative electron density; blue, positive electron density; σ , r.m.s. electron density for the unit cell. Significant structural

rearrangements are seen to extend from the active site to the extracellular surface in the vicinity of the putative proton translocation path. Figures were drawn using a modified version of MOLSCRIPT²⁹ and were rendered using Raster 3D³⁰.

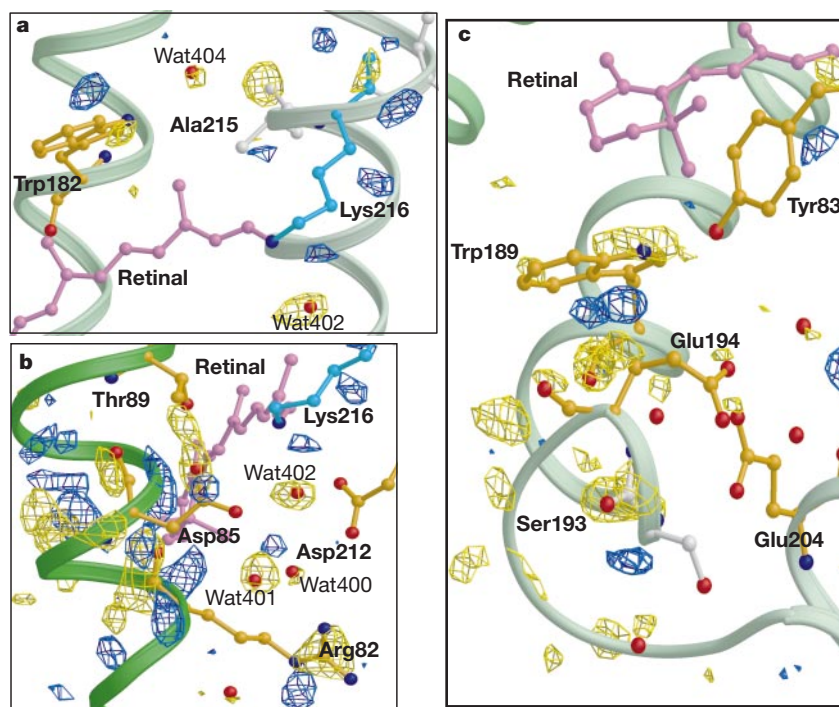


Figure 3 Detailed view of electron density changes overlaid on the ground state model and contoured to 3.4σ . **a**, View of the cytoplasmic side of the active site showing the retinal (purple) and negative and positive peaks near Lys 216 (blue), Ala 215 (white) and Trp 182 (orange). **b**, View showing paired positive and negative density features from residues 81 to 89, illustrating a bend in helix C centred on Asp 85 (orange) and a movement of Arg 82 (orange). Negative densities arise on Wat 402, Wat 401 and Wat 400

(waters are shown in red), and a positive peak appears at their centre. **c**, View showing electron density changes near the extracellular surface. Movements of Tyr 83 (orange), Trp 189 (orange) and Ser 193 (white) are apparent, and partial disordering is indicated by negative densities near the F–G loop. Yellow, negative electron density; blue, positive electron density.

density changes ($F_{\text{exc}} - F_{\text{ground}}$) overlaid on the ground state structure¹⁰. All significant changes associated with the earlier K_{IT} intermediate² are visible, with both positive and negative electron density peaks appearing along the side chain and backbone of Lys 216, near Asp 85, and a strong negative density peak showing the dislocation of water molecule Wat 402. These early movements, however, are propagated from the active site towards the extracellular surface and a considerable number of new structural rearrangements appear. These changes were not observed in electron diffraction studies in projection at 3.5 Å resolution⁸.

Figure 3a illustrates that the early perturbation of the hydrogen-bond network of helix G centred on Lys 216, observed in K_{IT} , has evolved to the cytoplasmic side of the active site, with negative and positive electron density peaks appearing on Ala 215. These changes represent the initial stages of a deformation² that facilitates significant movements of helices F and G later in the photocycle^{11,12}, and ultimately renders the Schiff base accessible to the cytoplasmic side. A positive density peak between the carbonyls of Lys 216 and Ala 215 indicates the appearance of a water molecule, which could facilitate the later reprotonation of the Schiff base from Asp 96, consistent with the Fourier transform infrared spectroscopy (FTIR) observation that the backbone carbonyl group of Lys 216 becomes solvated during the photocycle¹³. As anticipated¹¹, an upward movement of the indole ring of Trp 182 appears as a consequence of retinal isomerisation, and a hydrogen bond from Trp 182 to Wat 404 is maintained by the translation of this water molecule. In contrast to the late M state of the D96N mutant of bR¹⁴, the cytoplasmic ends of helices F and G are well ordered in L_{IT} .

The most striking structural rearrangements, however, occur within the extracellular half of the putative proton translocation pathway (Fig. 3b). In the ground state structure, an extended hydrogen-bonded network of water molecules and polar residues links helices C to G^{10,15}. The dislocation of Wat 402 is a key event for the K_{IT} intermediate² immediately following retinal isomerization. Consequent to this dislocation, negative electron density peaks arise on both Wat 401 (partially disordered in K_{IT}) and Wat 400, and a positive peak appears, centred between these three negative peaks, corresponding to the relocation of one of these water molecules¹⁴. The disruption of the ground-state network of water molecules allows the movement of helix C, which bends towards the putative proton translocation pathway (visualized as a sequence of paired negative and positive electron density peaks from residues 81 to 89, Fig. 3b). Furthermore, this disruption breaks the coupling¹⁶ of the

protonation state of Asp 85 to the pKa of the (putative) internal group, which releases a proton to the extracellular medium. Decoupling before the L-to-M transition was anticipated from kinetic studies¹⁷. As a consequence of this decoupling and the breakage of its hydrogen bond to Thr 89 (negative density between Oδ1 of Asp 85 and Oγ of Thr 89, Fig. 3b), the affinity of Asp 85 for protons is increased¹⁸, favouring proton transfer from the Schiff base to Asp 85. Paired negative/positive features in Fig. 3b reveal a flip in the orientation of Arg 82, approaching Oε1 of Glu 194 and favouring the release of a proton to the extracellular medium^{14,19}.

Significant structural rearrangements further assisting proton release are evident on the reverse side of helix C. A translation of Trp 86 within the hydrophobic retinal-binding pocket towards the ionone ring (not shown) and a movement of the side chain of Tyr 83 (helix C, Fig. 3c) are observed. Tyr 83 remains hydrogen-bonded to Trp 189 (helix F), which is then pushed toward the extracellular side, pressing on Glu 194 (Fig. 3c). This initiates a rearrangement of residues 190 to 195 in the F–G loop, partially exposing these residues to the extracellular medium and anticipating the proton release¹⁹.

High resolution X-ray structures of the ground state of bacteriorhodopsin^{10,15} reveal an asymmetric distribution of temperature factors along all seven helices. The cytoplasmic side of the protein displays local flexibility, with its structure being maintained through a tight packing of helices, whereas the extracellular side represents a stiff but fragile structure maintained by an extended hydrogen-bond network involving water molecules. The early K_{IT} intermediate structure² identifies the first dislocation of a key water molecule (Wat 402) separating the positively charged Schiff base and the negatively charged Asp 85 and Asp 212. As we observed here, this event triggers the collapse of the entire water-mediated hydrogen-bonded network from the Schiff base to Arg 82. This facilitates a flexing of helix C (Figs 3b, 4 and 5) that exaggerates the early movement² of Asp 85. Consequently Oδ2 of Asp 85 moves closer to Asp 212, and to the protonated Schiff base nitrogen, as suggested by NMR studies²⁰ showing an enhanced interaction between these two groups in the L-intermediate state.

Before proton transfer, the mutual electrostatic attraction of Asp 85 and the Schiff base³ drives, at least in part, the observed bend of helix C. Following the primary proton transfer event that neutralizes both groups, helix C could straighten out like an archer's bow and thereby inhibit the reprotonation of the Schiff base by Asp 85. Indeed, the structure of the deprotonated M-intermediate state of the D96N mutant of bacteriorhodopsin showed disordering of Wat 400, Wat 402, a translation of Wat 401 and the movement of Arg 82, but no displacement of the main chain of helix C was evident¹⁴, and studies of the kinetics of the L-to-M transition¹⁷ suggested that two protein conformations mediate the primary proton transfer equilibrium. Furthermore, efficient proton transfer

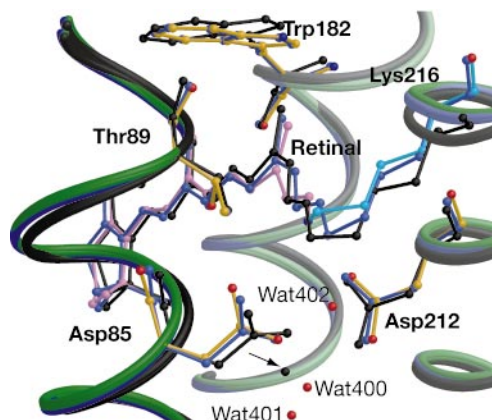


Figure 4 Structural rearrangements facilitating the primary proton transfer event. The refined model for the L_{IT} intermediate (black, Table 1) is superimposed upon the model for the K_{IT} intermediate² (dark blue, waters are omitted) and upon the ground state model (coloured as in Fig. 3). The reordered water in the L_{IT} model is highlighted by an arrow. The disruption of an extended hydrogen-bonded network of water molecules and polar residues facilitates a bend in helix C, which is also driven by the mutual electrostatic attraction of the Schiff base and Asp 85.

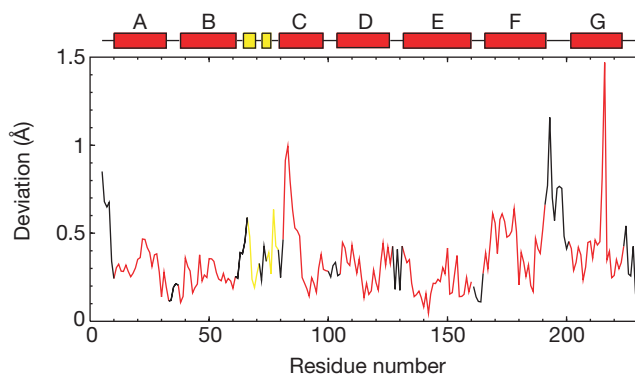


Figure 5 The r.m.s. deviation between the backbones of the refined L_{IT} model (Table 1) and the ground state bacteriorhodopsin versus the residue number. Red, α -helices; yellow, antiparallel β -sheet; black, loops.

occurs only when residue 85 is charged²¹, and the D85E mutation²² induces a tenfold acceleration of the L-to-M transition, supporting a mechanism where an extended movement of Asp 85 is key in proton transfer. A complementary proposal, where a change in the curvature of retinal after deprotonation induces a switch in the accessibility of the Schiff base¹², also assigns vectoriality to a physical separation of the Schiff base and Asp 85 following the initial proton transfer event.

In combination with the structures of the ground state^{10,15} and the structural rearrangements visualized immediately following retinal isomerisation², the movements described here present a consistent mechanistic picture, in atomic detail, of the events starting with photoexcitation and leading to the primary proton transfer event of the bacteriorhodopsin photocycle. This picture establishes that the early dislocation of a single key water molecule facilitates a cascade of structural rearrangements propagating from the active site towards the extracellular surface, inducing the transient deformation of an α -helix. This deformation facilitates the primary proton transfer event and a relaxation of helix C would then perturb the L-to-M equilibrium to ensure the vectoriality of the proton transport mechanism. □

Methods

Protein crystallization

Detergent-solubilized bacteriorhodopsin was crystallized within the mono-olein-based lipidic cubic phase as described⁴. The crystallization matrix was liquefied with lipase²³ before crystal mounting.

Crystal illumination and single crystal microspectrophotometry

Hexagonal bacteriorhodopsin crystals were light adapted by exposure to bright light for several minutes, mounted on cryoloops and then flash-frozen in liquid nitrogen. Frozen crystals were held within a stream of dry nitrogen gas cooled to 170 K and, after a delay of a few minutes, were illuminated for 30 s by light from a 100-mW laser diode ($\lambda = 532$ nm) guided to the crystal through an optical fibre². The large hexagonal face was orientated perpendicularly to the beam. Following a 40-s delay without illumination the crystals were immersed in liquid nitrogen. Spectral changes within crystals were monitored using single crystal microspectrophotometry⁹ and typical spectral results are shown in Fig. 1. The positive spectral feature at 412 nm, interpreted as representing a small contribution from a non-physiological or early M deprotonated state, became considerably stronger at higher light flux. To optimize the experimental conditions and to avoid pitfalls associated with mixed populations, spectral and structural changes were correlated over a variety of illumination intensities.

X-ray data collection

X-ray diffraction data, $\lambda = 0.934$ Å, were collected at beam line ID14-EH1 of the European Synchrotron Radiation Facility (ESRF), Grenoble, France, using a MAR CCD X-ray detector (MARResearch). A high intermediate population was trapped using an identical procedure to that described above, and crystals were maintained in the dark at 110 K during data collection. The lifetime of crystals within the X-ray beam was sufficient to allow the collection of one complete data set per crystal, typically consisting of 120 frames of 1° oscillation and 30 s exposure per frame (Table 1).

Data processing and structural analysis

Data were processed using the HKL package²⁴, PROW²⁵ and the CCP4 suite²⁶. The degree of merohedral twinning was determined for each data set using Britton plots and Yeates statistics (for a review see ref. 27). Crystals exhibited a variety of twinning ratios and intensities were corrected before structural analysis. Phases from the ground state¹⁰ (Protein data base entry 1qhj) were used to calculate Q-weighted²⁵ ($F_{\text{exc}} - F_{\text{ground}}$) electron density difference maps. The consistent appearance of (equally strong) correlated positive and negative features in the map (disordered water molecules excepted) indicated a low level of molecular disordering, such that a single excited state structural conformation could be assigned. As electron density difference maps display less model bias than the refined structure, they were used extensively to assign structural movements. The features described here were reproducibly observed in several illuminated data sets.

Crystallographic refinement

The L_{IT}-state population was initially estimated using a method derived from the Britton plot (E.P.-P. *et al.*, manuscript in preparation). The content of 70% L_{IT}-state thus obtained, which is consistent with population estimates⁸ of not less than 65% L from FTIR studies on purple membrane when illuminated at 170 K, was confirmed by occupancy refinement. The starting model contained two alternate conformations. Conformer 1 consisted of the ground state model 1qhj with the water molecules but without lipid molecules, given an occupancy of 0.30. Conformer 2 was set as 1qhj without the lipid and the water molecules, given an occupancy of 0.70. The models were first orientated using a rigid body alignment.

Simulated annealing followed by several cycles of energy minimization were then performed with CNS²⁸ on conformer 2 maintaining conformer 1 fixed. Water molecules observed in the difference maps were progressively added to conformer 2 during the refinement. Refinement statistics are shown in Table 1. Figure 5 shows the r.m.s. deviation of atomic positions and highlights the regions in which the differences between L_{IT} and ground states are the most pronounced. These are in accordance with the experimental Fo–Fo density maps (Figs 2 and 3).

Received 31 January; accepted 31 May 2000.

- Oesterhelt, D. & Stoekenius, W. Rhodopsin-like protein from the purple membrane of *Halobacterium halobium*. *Nature New Biol.* **233**, 149–152 (1971).
- Edman, K. *et al.* High-resolution X-ray structure of an early intermediate in the bacteriorhodopsin photocycle. *Nature* **401**, 822–826 (1999).
- Braiman, M. S. *et al.* Vibrational spectroscopy of bacteriorhodopsin mutants: Light-driven proton transport involves protonation changes of aspartic acid residues 85, 96 and 212. *Biochemistry* **27**, 8516–8520 (1988).
- Landau, E. M. & Rosenbusch, J. P. Lipidic cubic phases: A novel concept for the crystallization of membrane proteins. *Proc. Natl Acad. Sci. USA* **93**, 14532–14535 (1996).
- Schlichting, I., Berendzen, J., Phillips, G. N. & Sweet, R. M. Crystal structure of photolysed carbonmonoxymyoglobin. *Nature* **371**, 808–812 (1994).
- Genick, U. K., Soltis, S. M., Kuhn, P., Canestrelli, I. L. & Getzoff, E. D. Structure at 0.85 Å resolution of an early protein photocycle intermediate. *Nature* **392**, 206–209 (1998).
- Maeda, A., Sasaki, J., Yamazaki, Y., Needleman, R. & Lanyi, J. K. Interaction of Aspartate-85 with a water molecule and the protonated Schiff base in the L intermediate of bacteriorhodopsin: A Fourier-transform infrared spectroscopy study. *Biochemistry*, **33**, 1713–1717 (1994).
- Hendrickson, F. M., Burkard, F. & Glaeser, R. M. Structural characterization of the L-to-M transition of the bacteriorhodopsin photocycle. *Biophys. J.* **75**, 1446–1454 (1998).
- Hadfield, A. & Hajdu, J. A fast and portable microspectrophotometer for protein crystallography. *J. Appl. Crystallogr.* **26**, 839–842 (1993).
- Belrhali, H. *et al.* Protein, lipid and water organization in bacteriorhodopsin crystals: a molecular view of the purple membrane at 1.9 Å resolution. *Structure* **7**, 909–917 (1999).
- Subramaniam, S., Gerstein, M., Oesterhelt, D. & Henderson, R. Electron diffraction analysis of structural changes in the photocycle of bacteriorhodopsin. *EMBO J.* **12**, 1–8 (1993).
- Subramaniam, S. *et al.* Protein conformational changes in the bacteriorhodopsin photocycle. *J. Mol. Biol.* **287**, 145–161 (1999).
- Takei, H., Gat, Y., Rothman, Z., Lewis, A. & Sheves, M. Active site Lysine backbone undergoes conformational changes in the bacteriorhodopsin photocycle. *J. Biol. Chem.* **269**, 7387–7389 (1994).
- Luecke, H., Schobert, B., Richter, H. T., Cartailler, J.-P. & Lanyi, J. K. Structural changes in bacteriorhodopsin during ion transport at 2 Å resolution. *Science* **286**, 255–260 (1999).
- Luecke, H., Schobert, B., Richter, H. T., Cartailler, J.-P. & Lanyi, J. K. Structure of bacteriorhodopsin at 1.55 Å resolution. *J. Mol. Biol.* **291**, 899–911 (1999).
- Balashov, S. P., Imasheva, E. S., Govindjee, R. & Ebrey, T. G. Titration of aspartate-85 in bacteriorhodopsin: what it says about chromophore isomerization and proton release. *Biophys. J.* **70**, 473–481 (1996).
- Althaus, T. & Stockburger, M. Time and pH dependence of the L-to-M transition in the photocycle of bacteriorhodopsin and its correlation with proton release. *Biochemistry* **37**, 2807–2817 (1998).
- Gat, Y. & Sheves, M. A mechanism for controlling the pKa of the retinal protonated Schiff base in retinal proteins. A study with model compounds. *J. Am. Chem. Soc.* **115**, 3772–3773 (1993).
- Balashov, S. P. *et al.* Glutamate-194 to cysteine mutation inhibits fast light-induced proton release in bacteriorhodopsin. *Biochemistry* **36**, 8671–8676 (1997).
- Hu, J. G., Sun, B. Q., Petkova, A. T., Griffin, R. G. & Herzfeld, J. The predischarge chromophore in bacteriorhodopsin: a ¹⁵N solid state NMR study of the L photointermediate. *Biochemistry* **36**, 9316–9322 (1997).
- Khorana, H. G. Two light-transducing membrane proteins: bacteriorhodopsin and the mammalian rhodopsin. *Proc. Natl Acad. Sci. USA* **90**, 1166–1171 (1993).
- Heberle, J., Oesterhelt, D. & Dencher, N. A. Decoupling of photo- and proton cycle in the Asp → Glu mutant of bacteriorhodopsin. *EMBO J.* **12**, 3721–3727 (1993).
- Nollert, P. & Landau, E. M. Enzymic release of crystals from lipidic cubic phases. *Biochem. Soc. Trans.* **26**, 709–713 (1998).
- Otwinski, Z. In *Data Collection and Processing* (eds Sawyer, L., Isaacs, N. W. & Bailey, S.) 55–62 (DL/SCI/R34, Daresbury Laboratory, Warrington, 1993).
- Bourgeois, D. New processing tools for weak and/or spatially overlapped macromolecular diffraction patterns. *Acta Crystallogr. D* **55**, 1733–1741 (1999).
- Collaborative computational project nr4. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Chandra, N., Ravi Acharya, K. & Moody, P. C. E. Analysis and characterization of data from twinned crystals. *Acta Crystallogr. D* **55**, 1750–1758 (1999).
- Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Esnouf, R. M. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 133–138 (1997).
- Merrit, E. A. & Murphy, M. E. P. Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873 (1994).

Acknowledgements

We thank H. Belrhali, D. Bourgeois, J. Hajdu, A. Hardmeyer, J. Navarro, P. Nollert, H. Pettersson and S. Ramaswamy for experimental contributions and discussions, and G. Büldt for providing purple membrane. Support from the EU-BIOTECH, the Swedish Natural Science Research Council (NFR) and the Swiss National Science Foundation's SPP BIOTECH is gratefully acknowledged.

Correspondence and requests for materials should be addressed to E.M.L. (e-mail: emlandau@utmb.edu). Crystallographic co-ordinates and structure factor amplitudes have been deposited with the Protein Data Bank (accession code: 1EOP).

Structural alterations for proton translocation in the M state of wild-type bacteriorhodopsin

Hans Jürgen Sass*, Georg Büldt*, Ralf Gessenich*, Dominic Hehn*, Dirk Neff*, Ramona Schlesinger*, Joel Berendzen† & Pal Ormos‡

* Institute of Structural Biology, Research Centre Jülich, 52425 Jülich, Germany

† Biophysics Group, Mail Stop D454, Los Alamos Laboratory, Los Alamos, New Mexico 87545, USA

‡ Institute of Biophysics, Biological Research Centre of the Hungarian Academy of Sciences, Szeged, PO Box 521, 6701 Szeged, Hungary

The transport of protons across membranes is an important process in cellular bioenergetics. The light-driven proton pump bacteriorhodopsin is the best-characterized protein providing this function. Photon energy is absorbed by the chromophore retinal, covalently bound to Lys 216 via a protonated Schiff base. The light-induced all-*trans* to 13-*cis* isomerization of the retinal results in deprotonation of the Schiff base followed by alterations in protonatable groups within bacteriorhodopsin. The changed force field induces changes, even in the tertiary structure^{1–3}, which are necessary for proton pumping. The recent report⁴ of a high-resolution X-ray crystal structure for the late M intermediate of a mutant bacteriorhodopsin (with Asp 96→Asn) displays the structure of a proton pathway highly disturbed by the mutation. To observe an unperturbed proton pathway, we determined the structure of the late M intermediate of wild-type bacteriorhodopsin (2.25 Å resolution). The cytoplasmic side of our M₂ structure shows a water net that allows proton transfer from the proton donor group Asp 96 towards the Schiff base. An enlarged cavity system above Asp 96 is observed, which facilitates the de- and reprotonation of this group by fluctuating water molecules in the last part of the cycle.

The changes in the interaction of the protein matrix with the chromophore allow the identification of the main functional states of bacteriorhodopsin: namely, the light-adapted ground state (bR₅₆₈), and the photocycle intermediates K₅₉₀, L₅₅₀, M₁₄₁₂ (early M), M₂₄₁₂ (late M), N₅₃₀ and O₆₄₀ (Here subscripts give the absorption maxima in nm, and we use bR as an abbreviation for the light-adapted ground state of bacteriorhodopsin.) The extent of conformational changes and their connection with the pumping process has become clear through the investigation of the M state, the key intermediate of the photocycle (refs 1–4). Between the early and late M states, the accessibility of the Schiff base for protons has switched. In the first part of the photocycle up to M₁, the Schiff base is connected to the extracellular side. After M₂ the Schiff base can only be reprotonated from the cytoplasmic side by Asp 96 with the transition to the following N intermediate^{5,6}. It is this part of the proton transport process which is disturbed by the mutation Asp 96→Asn (refs 7, 8).

In response to the absorption of a photon, the retinal and its immediate environment are changed during intermediates K and L to allow the proton transfer from the Schiff base to Asp 85. According to the locations and movements of water molecules displayed in Figs 1 and 2a, 3b, we suggest that this occurs via water molecule W710_{bR} located in the bR structure between the Schiff base, Asp 85 and Asp 212. This water molecule, which corresponds to W402_{bR} (abbreviation for W402 in Asp 96→Asn ground-state structure⁹), is shifted in both M-state structures by 2.6 Å to the extracellular side (Fig. 4). Curiously, in the K-state structure¹⁰ the corresponding W710_{bR} has vanished, so that in line with Fourier transform infrared (FTIR) spectroscopy, we have to assume¹¹ that it reappears during the L state in this region. Asp 85

displays a similar orientation in K as in our M state, so this conformation seems to be related to the presence or absence of this water molecule. Furthermore, we have to assume that this configuration is also restored or even optimized for the proton translocation following the L intermediate. The proton transfer from the Schiff base towards Asp 85 initiates the large downward reorientation of the positively charged Arg 82 side chain towards the dyad Glu 194/Glu 204. A switch function of Arg 82 was predicted from theoretical considerations¹². The strong new interaction of Arg 82 with the dyad leads to a conformational change in Glu 204 and a larger separation between the glutamic acid residues.

We find that the extent of reorientation of Arg 82 and Glu 204 is much more pronounced in our wild-type bacteriorhodopsin than in the mutant structure⁴ (Fig. 4). This change in the charge distribution and in the hydrogen bonding clearly affects the pK_a values of the dyad so that a proton is released either directly from one of the glutamic acids¹³ or indirectly via a net of water molecules^{14,15}. The strong hydrogen-bond interaction of Arg 82 NH₂-groups with the Glu 204 carboxyl group in our M-state model, absent in the mutant M structure, reinforces the proposal of Richter *et al.*¹⁶ that Glu 204 is the proton-release group. The strong hydrogen-bond interaction may explain why no pronounced difference band was observed in FTIR spectra¹⁵.

The 13-*cis* isomerization of the retinal with the Schiff-base nitrogen oriented upwards towards the Thr 89 hydroxyl group, the reorientation of Asp 85, the downward displacement of W710, and the reorientation of Arg 82 together with associated water molecules results in the isolation of the Schiff-base nitrogen away from the new network of hydrogen bonds in the extracellular domain (Fig. 3b). This is an important element to establish the

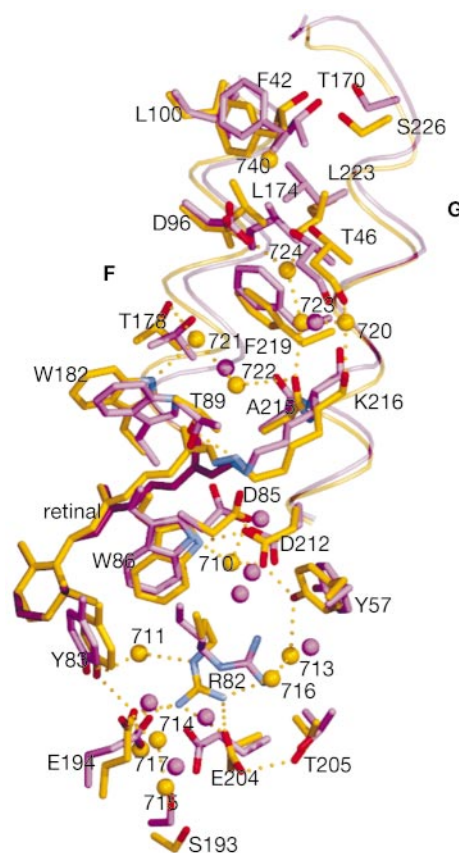


Figure 1 Image of the proton pathway of the BR model (purple) and the M₂-state model (yellow). Shown are several functionally important amino acids, the retinal, water molecules and the cytoplasmic parts of helices F and G, showing the most important alterations between these photocycle states. Top, cytoplasmic side.

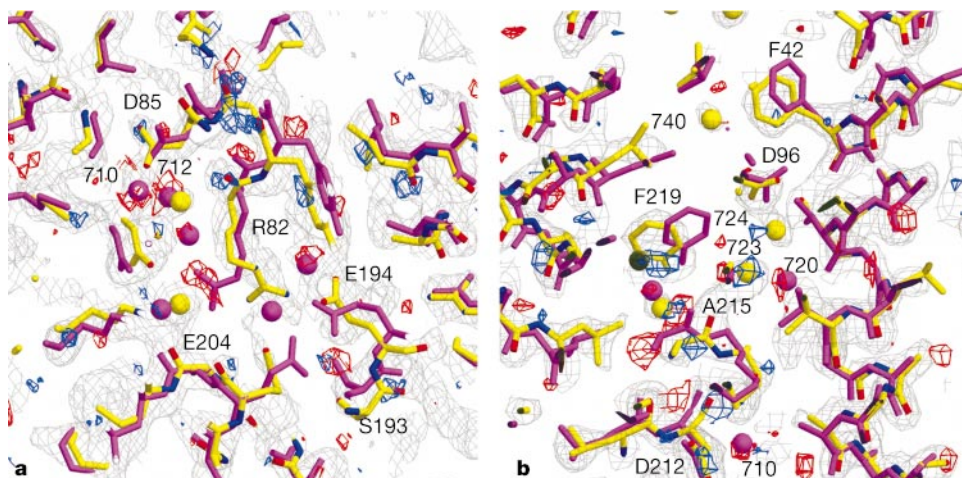


Figure 2 Details of the structural differences between the ground state and the M2 intermediate. **a**, For the extracellular part of the molecule, below the retinal; **b**, for the cytoplasmic part, above the retinal. In both images the positive Fourier difference densities are displayed in blue, indicating atoms moved to this position in the M2 state and negative

difference densities in red, indicating atoms removed from this position. The electron densities displayed in grey are in **a** extrapolated omit densities (see Methods) contoured at 0.6σ level and in **b** are M select densities $\Delta\rho_s$ at the 1σ level. The model of the M2 structure is represented in yellow and the bR model in purple.

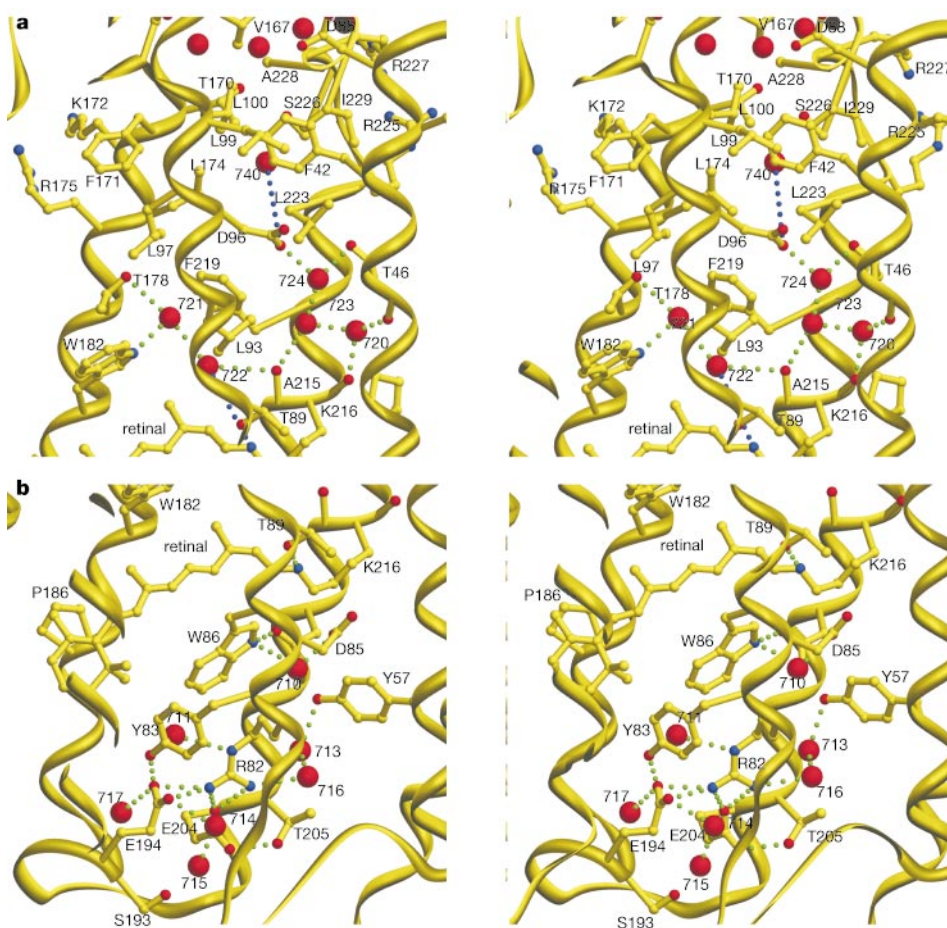


Figure 3 Stereo images of the M2-state model of wild-type bacteriorhodopsin. **a**, The cytoplasmic part; **b**, the extracellular part. Green dotted lines are hydrogen bonds in the range between 2.4 and 3.7 Å. Water molecules are displayed as red balls. Panel **a** shows the water network from Asp 96 to the carbonyl oxygen of Ala 215, and the water molecule 740 in an intermediate position between the cytoplasmic surface and Asp 96. Distances

of about 5 to 6 Å in blue dotted lines indicate necessary movements of water molecules to complete the proton transport. Panel **b** displays the interactions of Arg 82, Glu 194 and Glu 204 with each other and with surrounding water molecules, as well as the possible interaction of Thr 89 with the Schiff base of the 13-*cis* retinal.

vectoriality of the pump, in addition to the marked changes of the pK_a -values⁵ of the Schiff base and Asp 85 between the bR and M states.

Considering the structural changes in the cytoplasmic part, as detected by low-resolution diffraction methods^{1–3}, it became evident that the extent of hydration¹⁷ and the redistribution of charges during proton translocation have a strong influence on the tertiary structure¹⁸. At high resolution, our late M-state model confirms the changes in the tertiary structure of bacteriorhodopsin; these changes consist of a shift of helix F in the region from Trp 189 to Trp 182 by about 0.8 Å towards the cytoplasmic side and an outward tilt in the region from Val 179 to Val 167 (Fig. 1). This tilt is not homogeneous along F; rather, it is more a bending (as proposed from molecular dynamic studies¹⁹), with the largest bending (about 1.3 Å at Cα positions) occurring from Leu 174 to Asn 176. Changes in helix G can be described as a narrowing of the helix turn in the region of Leu 221 to Arg 227 and as an opening and an increase in curvature near Ala 215 and Lys 216 (Figs 1, 3a). The M-state model of the mutant⁴ can only partly display these alterations because the upper parts of helices F and G are missing.

In the bR state, helices F and G are bridged via W721_{bR}, which is connecting Nε of Trp 182 to the carbonyl Ala 215 (Fig. 1), adopting already in the ground state an unusual π -helix conformation in this region⁹. In our M-state model this bridge is broken, and W721_M connects the Nε of the upwardly displaced Trp 182 to the hydroxyl group of Thr 178. The direct interaction of W721 with the carbonyl oxygen of Ala 215 is no longer possible. However, this configuration allows the stabilization of the additional water molecule, W722_M, between W721_M and carbonyl oxygen Ala 215 (Figs 1, 3a). The greater flexibility of W722_M, indicated by the extended hydrogen-bond distance of about 3.7 Å to the oxygen of Ala 215 and the higher

B-value, could be important for the Schiff-base reprotonation. Thus in the M state, water molecule W721 together with the carbonyl group of Ala 215 have located W722_M in close proximity for Schiff-base reprotonation.

The changes in the tertiary structure, together with the reorientation of side chains—especially the movement of Phe 219 (Figs 1, 2b, 3a)—have enlarged several cavities in the cytoplasmic domain (see Supplementary Information) which are important for the reprotonation of the Schiff base and Asp 96. An upper cavity is connected to the cytoplasmic surface. The alterations also allow new hydrogen bonds in the cavity between Asp 96 and Lys 216. In the bR state, the carbonyl oxygen 216 forms a helical hydrogen bond to the amide proton Gly 220 and to W720_{bR} which is also stabilized by a hydrogen bond to the carbonyl oxygen of Thr 46 in helix B. This is possible as the helical hydrogen bond of Thr 46 is broken owing to Pro 50. The interaction of water molecule W720 with the carbonyl oxygen 216 persists in the M₂ structure (Figs 1, 3a). Consequently, we believe it is owing to this interaction that the 13-*cis* retinal exerts an outward bending force on the local region of helix G via a conformational change of the Lys 216 side chain (Figs 1, 3a). This results in the disruption of the hydrogen bond to Gly 220 and an outward tilt of the 216 carbonyl.

The relocation of W720 allows two further water molecules, W723_M and W724_M, to appear in the M₂-state structure (Figs 1, 3a). W724_M interacts with Oδ₁ of Asp 96 and Oγ₁ of Thr 46 which are in hydrogen-bond contact in bR. W723_M is strongly fixed in its position owing to hydrogen-bond contacts with W724_M, W720_M and the carbonyl oxygen of Ala 215. This contradicts the results for the Asp 96→Asn mutant, but is in agreement with FTIR measurements²⁰ which suggested a water molecule in the region near the Schiff base, Phe 219 and Val 49. We suggest that W723_M may be absent owing to the mutation, but is of special importance for proton transport through the cytoplasmic part. Together these water molecules form a proton transfer chain from Asp 96 → W724_M → W723_M → O-Ala 215 → W722_M → Schiff base. Only the remaining gap (of 5.8 Å) between W722_M and the Schiff base is too large for a direct proton transfer (blue dotted line in Fig. 3a). We suggest that this water molecule, as it is less stabilized than the nearest neighbour water molecules and because there is no further barrier in between, can fluctuate to the Schiff base when the proton has reached the carbonyl oxygen region of Ala 215. This hydrogen-bond network above the Schiff base is an important new feature of our late M-state structure, and is essential for Schiff base reprotonation.

The pathway from the cytoplasmic surface towards Asp 96 is important not only for the reprotonation of this amino-acid residue, but also for the initiation of the Schiff-base reprotonation. According to the structure, the transfer through this part is very different from proton translocation processes in other parts of the molecule. This structural domain can not be compared with the mutant M-state model which does not contain amino acids 154 to 175 and 223 to 248. In this part of our M-state structure, an electron density in the larger cavity formed in the M₂ state was attributed to a crystallographically unusual water molecule, W740_M, which is surrounded by mostly hydrophobic residues (Thr 170, Phe 42, Leu 100, Leu 223, Ile 229) without hydrogen-bonding partners in its vicinity. W740_M is at a distance of 5.35 Å from Asp 96. The cavity has a narrow opening to the cytoplasmic surface formed by the nonpolar parts of the residues (Ala 228, Thr 170, Leu 100, Ile 229, Ser 226). The higher flexibility and the partial opening in this part of the molecule would allow a water molecule to penetrate in a directed random walk and to be trapped in this hydrophobic cavity. This suggests that the contact to Asp 96 might be achieved in the millisecond time domain by a fluctuating water molecule and not by a continuous water channel. The interaction between W740_M and Asp 96 would lower the pK_a of this residue and in this way initiate the deprotonation of Asp 96 and thereafter stabilize the

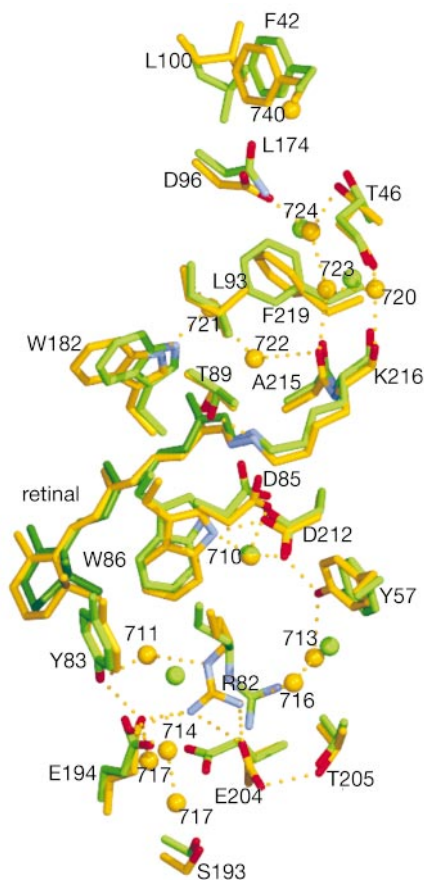


Figure 4 Comparison of the late M-state models of wild-type bacteriorhodopsin (yellow) with the mutant Asp96→Asn (green; ref. 4). Shown are several functionally important amino acids, the retinal, and water molecules. Top, cytoplasmic side.

charged state of this group. An aqueous channel was also excluded from low-resolution neutron diffraction work²¹.

Asp 96 is located at a strategic position between two very different parts of the proton pathway. One consists of a stabilized chain of water molecules directed towards the Schiff base; the other, directed towards the cytoplasmic surface, is confined by hydrophobic amino acids surrounding a fluctuating water molecule. The different conductivities for protons of these two pathways establish a vectorial proton transport to the Schiff base, as displayed in our M-state structure. In the subsequent reaction, the reprotonation of Asp 96 may also occur by means of a fluctuating water molecule from the cytoplasmic surface. □

Methods

Specimens

Crystals were grown according to the protocol of Landau and Rosenbusch²². Diffraction experiments at 95 K were performed at the X8C beam line of the NSLS Brookhaven with a wavelength $\lambda = 1.214 \text{ \AA}$.

Formation of M intermediate

The crystal was heated to room temperature (by blocking the nitrogen cold stream), illuminated with green light (33 mW, 514 nm) for 1 s, then cooled (by unblocking the cold stream) while the green light was still on. One second after recooling started, the illumination was turned off.

Microspectroscopy of the crystal

The amount of components in the photoproduct after illumination was estimated on the basis of known spectra of the intermediates²³. On average 60–70% of the bacteriorhodopsin molecules in each crystal were accumulated in the M state, small amounts in the N intermediate, and 30–40% in the bR state.

Treatment of diffraction data

Crystallographic parameters of the crystals with the space group $P6_3$ ($\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$) were determined for the ground-state crystal to be $a^{\text{bR}} = b^{\text{bR}} = 60.81 \text{ \AA}$, $c^{\text{bR}} = 110.64 \text{ \AA}$ ($R_{\text{sym}} = 4.7\%$, $20\text{--}2.0 \text{ \AA}$), and for the illuminated crystal $a^{\text{I}} = b^{\text{I}} = 61.08 \text{ \AA}$, $c^{\text{I}} = 110.40 \text{ \AA}$ ($R_{\text{sym}} = 8.2\%$, $20\text{--}2.25 \text{ \AA}$). CNS²⁴ version 0.4 was used for most of the further processing of the data. The twinning ratio was first determined by using the twinning server at UCLA (www.doe-mbi.ucla.edu/Services/Twinning/), Yeates, T.O.) and then refined. The final twinning ratio was 0.44 for the ground-state crystal and 0.38 for the illuminated one. To cope with the problem of twinning, the CNS program that we used was slightly modified; for any comparison of measured amplitudes with model amplitudes, a twinned version of the model amplitudes was used.

Ground-state refinement

As a starting model, that of Luecke *et al.*²⁵, PDB:1BRX, was used, and refined to final R values of $R = 17.9\%$ and $R_{\text{free}} = 20.8\%$.

M-state refinement

The bR model was first oriented in the new unit cell with a molecular replacement calculation and a rigid body refinement. Six different approaches were used for cross-checking the molecular alterations found for the M2 intermediate. (1) Fourier difference density maps ($F_o^{\text{I}} - F_o^{\text{bR}}$) of the observed structure factors of the illuminated and the ground-state crystal) at different resolutions were calculated, and used to locate the changes in the molecule. (2) A model of two complete bacteriorhodopsin molecules including water molecules and hydrocarbon chains, one declared as M and one as bR, was refined against the data. During the energy minimization refinement, the atomic positions of the bR model were fixed whereas the atomic coordinates of the M model with the complementary occupancy were allowed to change. In the same way, simulated annealing calculations (with standard slow cool protocol at 2,500 K and 4,000 K) were performed, which resulted in M-state models comparable to that obtained from simple energy minimization. (3) The observed structure factors F_o^{I} of the illuminated crystal contain a fraction s of the bR state and a fraction $1-s$ of the M state. Therefore, extrapolated difference density maps²⁶ $\Delta\rho_{\text{ex}} = (|F_o^{\text{I}}| - s|F_o^{\text{bR}}|)[\exp(i\varphi_c^{\text{I}})/(1-s)]$ were calculated using amplitudes F_o^{bR} and phases φ_c^{bR} of the ground-state model. These maps should show the electron density of the M intermediate for the correct fraction s of the ground state contribution to F_o^{I} . (4) Omit density maps²⁷ $\Delta\rho_{\text{exo}}$ were calculated by the $\Delta\rho_{\text{ex}}$ formula using phases φ_c^{bR} determined without including the retinal and/or Arg 82, Asp 85, Glu 194, Glu 204, Lys 216, Phe 219 and all water molecules (Fig. 2a). These maps were used for visual comparison of the M-state conformations of these residues with the resulting densities. (5) Complete omit density maps $\Delta\rho_o$ were calculated with φ_c^{bR} as well as φ_c^{bR} determined from a bR model without the above-mentioned amino acids and/or the retinal. Varying fractions of the M and bR state conformations of the omitted residues were correlated to the $\Delta\rho_o$ densities to define the fraction of the M conformation. The correlations resulted in about 65% 13-*cis* retinal and only about 35% M-state conformation for the respective amino-acid side chains. If we consider that the M intermediate consists of the early and the late M state contributions and assume that the early M-state structure of these residues are very similar

to their bR state conformations as verified at low resolution, we conclude that on average 35% of bacteriorhodopsin molecules in an illuminated crystal are in the ground state, another 35% in the late M state and 30% in the early M. (6) Selected difference density maps $\Delta\rho_s = (|F_o^{\text{I}}| - 0.65|F_o^{\text{bR}}|)[\exp(i\varphi_c^{\text{I}})/(1 - 0.65)]$ were calculated using amplitudes F_o^{bR} of the ground-state model and phases φ_c^{I} of the M intermediate model (Fig. 2b). These maps show the final electron density of the M intermediate. The final R values of the M state model, including the fixed ground-state model, are $R = 16.7\%$ and $R_{\text{free}} = 23.6\%$, corresponding to an estimated coordinate error using R_{free} of $\Delta r = 0.279 \text{ \AA}$, from the Luzzati plot, and $\Delta r = 0.239 \text{ \AA}$, from the SIGMAA analysis, as determined with standard input-files of CNS. Figures 2 and 3 were prepared with the program²⁸ SETRO, and Figs 1 and 4 with POV-RAY (POV-Ray team: <http://www.povray.org>). A more complete description of the technical details, and tables with crystallographic information, are given in Supplementary Information.

Received 26 November 1999; accepted 18 May 2000.

- Dencher, N. A., Dresselhaus, D., Zaccai, G. & Büldt, G. Structural changes in bacteriorhodopsin during proton translocation revealed by neutron diffraction. *Proc. Natl Acad. Sci. USA* **86**, 7876–7879 (1989).
- Koch, M. H. J. *et al.* Time-resolved X-ray diffraction study of structural changes associated with the photocycle of bacteriorhodopsin. *EMBO J.* **10**, 521–526 (1991).
- Subramaniam, S., Gerstein, M., Oesterheld, D. & Henderson, R. Electron diffraction analysis of structural changes in the photocycle of bacteriorhodopsin. *EMBO J.* **12**, 1–8 (1993).
- Luecke, H., Schober, B., Richter, H. T., Cartailler, J. P. & Lanyi, J. K. Structural changes in bacteriorhodopsin during ion transport at 2 angstrom resolution. *Science* **286**, 255–261 (1999).
- Lanyi, J. K. Understanding structure and function in the light-driven proton pump bacteriorhodopsin. *J. Struct. Biol.* **124**, 164–178 (1998).
- Oesterheld, D. The structure and mechanism of the family of retinal proteins from halophilic archaea. *Curr. Opin. Struct. Biol.* **8**, 489–500 (1998).
- Butt, H. J., Fendler, K., Bamberg, E., Tittor, J. & Oesterheld, D. Aspartic acids 96 and 85 play a central role in the function of bacteriorhodopsin as a proton pump. *EMBO J.* **8**, 1657–1663 (1989).
- Gerwert, K., Hess, B., Soppa, J. & Oesterheld, D. Role of aspartate-96 in proton translocation by bacteriorhodopsin. *Proc. Natl Acad. Sci. USA* **86**, 4943–4947 (1989).
- Luecke, H., Schober, B., Richter, H. T., Cartailler, J. P. & Lanyi, J. K. Structure of bacteriorhodopsin at 1.55 Å resolution. *J. Mol. Biol.* **291**, 899–911 (1999).
- Edman, K. *et al.* High-resolution X-ray structure of an early intermediate in the bacteriorhodopsin photocycle. *Nature* **401**, 822–826 (1999).
- Maeda, A., Sasaki, J., Yamazaki, Y., Needleman, R. & Lanyi, J. K. Interaction of aspartate-85 with a water molecule and the protonated Schiff base in the L intermediate of bacteriorhodopsin: a Fourier-transform infrared spectroscopic study. *Biochemistry* **33**, 1713–1717 (1994).
- Scharnagl, C. & Fischer, S. F. Conformational flexibility of arginine-82 as source for the heterogeneous and pH-dependent kinetics of the primary proton transfer step in the bacteriorhodopsin photocycle: An electrostatic model. *Chem. Phys.* **212**, 231–246 (1996).
- Dioumaev, A. K. *et al.* Existence of a proton transfer chain in bacteriorhodopsin: participation of Glu-194 in the release of protons to the extracellular surface. *Biochemistry* **37**, 2496–2506 (1998).
- Rammelsberg, R., Huhn, G., Lubbén, M. & Gerwert, K. Bacteriorhodopsin's intramolecular proton-release pathway consists of a hydrogen-bonded network. *Biochemistry* **37**, 5001–5009 (1998).
- Zscherp, C., Schlesinger, R., Tittor, J., Oesterheld, D. & Heberle, J. *In situ* determination of transient pKa changes of internal amino acids of bacteriorhodopsin by using time-resolved attenuated total reflection Fourier-transform infrared spectroscopy. *Proc. Natl Acad. Sci. USA* **96**, 5498–5503 (1999).
- Richter, H. T., Brown, L. S., Needleman, R. & Lanyi, J. K. A linkage of the pKa's of asp-85 and glu-204 forms part of the reprotonation switch of bacteriorhodopsin. *Biochemistry* **35**, 4054–4062 (1996).
- Sass, H. J. *et al.* The tertiary structural changes in bacteriorhodopsin occur between M states: X-ray diffraction and Fourier transform infrared spectroscopy. *EMBO J.* **16**, 1484–1491 (1997).
- Sass, H. J. *et al.* Evidence for charge-controlled conformational changes in the photocycle of bacteriorhodopsin. *Biophys. J.* **75**, 399–405 (1998).
- Xu, D., Sheves, M. & Schulten, K. Molecular dynamics study of the M412 intermediate of bacteriorhodopsin. *Biophys. J.* **69**, 2745–2760 (1995).
- Yamazaki, Y., Kandori, H., Needleman, R., Lanyi, J. K. & Maeda, A. Interaction of the protonated Schiff base with the peptide backbone of valine 49 and the intervening water molecule in the N photointermediate of bacteriorhodopsin. *Biochemistry* **37**, 1559–1564 (1998).
- Weik, M., Zaccai, G., Dencher, N. A., Oesterheld, D. & Hauss, T. Structure and hydration of the M-state of the bacteriorhodopsin mutant D96N studied by neutron diffraction. *J. Mol. Biol.* **275**, 625–634 (1998).
- Landau, E. M. & Rosenbusch, J. P. Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. *Proc. Natl Acad. Sci. USA* **93**, 14532–14535 (1996).
- Zimanyi, L. & Lanyi, J. K. Deriving the intermediate spectra and photocycle kinetics from time-resolved difference spectra of bacteriorhodopsin. The simpler case of the recombinant D96N protein. *Biophys. J.* **64**, 240–251 (1993).
- Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Luecke, H., Richter, H. T. & Lanyi, J. K. Proton transfer pathways in bacteriorhodopsin at 2.3 angstrom resolution. *Science* **280**, 1934–1937 (1998).
- Genick, U. K. *et al.* Structure of a protein photocycle intermediate by millisecond time-resolved crystallography. *Science* **275**, 1471–1475 (1997).
- Bhat, T. N. & Cohen, G. H. OMITMAP: An electron density map suitable for the examination of errors in a macromolecular model. *J. Appl. Crystallogr.* **17**, 244–248 (1984).
- Evans, S. V. SETOR: hardware-lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graph.* **11**, 134–138 (1993).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank E. Landau, J. Rosenbusch, J. Heberle, V. Gordeliy, J. Granzin and J. Labahn for support and discussions, and A. Cousin for preparation of purple membranes. This work was supported by EU-BIOTECH and the Deutsche Forschungsgemeinschaft, Sfb 189

(H.J.S. and G.B.), the US Department of Energy, Office of Health and Environmental Research (J.B.), and OTKA (P.O.).

Correspondence and requests for materials should be addressed to G. B. (e-mail: g.bueldt@fz-juelich.de). Coordinates have been deposited in the Protein Data Bank (accession code: 1CWQ).

Molecular mechanism of vectorial proton translocation by bacteriorhodopsin

Sriram Subramaniam*† & Richard Henderson*

*MRC Laboratory of Molecular Biology, Cambridge CB2 2QH, UK

†Laboratory of Biochemistry, National Cancer Institute, Bethesda, Maryland 20892, USA

Bacteriorhodopsin, a membrane protein with a relative molecular mass of 27,000, is a light driven pump which transports protons across the cell membrane of the halophilic organism *Halobacterium salinarum*. The chromophore retinal is covalently attached to the protein via a protonated Schiff base. Upon illumination, retinal is isomerized. The Schiff base then releases a proton to the extracellular medium, and is subsequently reprotonated from the cytoplasm. An atomic model for bacteriorhodopsin was first determined by Henderson *et al.*¹, and has been confirmed and extended by work in a number of laboratories in the last few years². Here we present an atomic model for structural changes involved in the vectorial, light-driven transport of protons by bacteriorhodopsin. A 'switch' mechanism ensures the vectorial nature of pumping. First, retinal unbends, triggered by loss of the Schiff base proton, and second, a protein conformational change occurs. This conformational change, which we have determined by electron crystallography at atomic (3.2 Å in-plane and 3.6 Å vertical) resolution, is largely localized to helices F and G, and provides an 'opening' of the protein to protons on the cytoplasmic side of the membrane.

A fundamental unanswered question regarding the mechanism of proton transport in bacteriorhodopsin is the structural nature of the 'switch' event, which ensures that the release and uptake of protons do not occur from the same side of the membrane. A protein conformational change that switches the accessibility of the active site is a very plausible candidate for the switching mechanism in any ion pump. Detailed analysis of conformational changes of bacteriorhodopsin at various times after illumination shows that a single, large protein conformational change to the 'cytoplasmically open' state occurs within 1 ms after illumination³. This change, which is roughly coincident with deprotonation of the Schiff base and formation of the M intermediate, persists through the rest of the photocycle, and is reversed upon the thermal regeneration of the starting protein conformation.

We merged electron diffraction patterns from 402 crystals of the D96G, F171C, F219L triple mutant of bacteriorhodopsin to obtain a set of structure factor amplitudes to 3.2 Å resolution and 62.5° specimen tilt for the cytoplasmically open state of bacteriorhodopsin. We chose this mutant because the full extent of the light-driven conformational change is present in this mutant³ in the presence or absence of illumination and in both fully hydrated membranes and membranes embedded in glucose or in trehalose.

Figure 1 shows two slices of the three-dimensional difference Fourier map between the D96G, F171C, F219L triple mutant and wild-type bacteriorhodopsin. Inspection of the map shows that the

largest features of the difference map are on the cytoplasmic side, in the vicinity of helices F and G. We used these features in the experimental difference map initially as a guide, and subsequently to confirm the results of the systematic refinement procedure used to determine an atomic model for the structure of the D96G, F171C, F219L triple mutant. In parallel, we also refined the structure of wild-type bacteriorhodopsin against electron diffraction data using the same protocol to produce a comparable electron crystallographic model for the protein in the unilluminated state.

Inspection of $2F_o - F_c$ maps (Fig. 2a and b) calculated from refined structures of wild-type bacteriorhodopsin and the D96G, F171C, F219L triple mutant and the superposed atomic models (Fig. 2c)

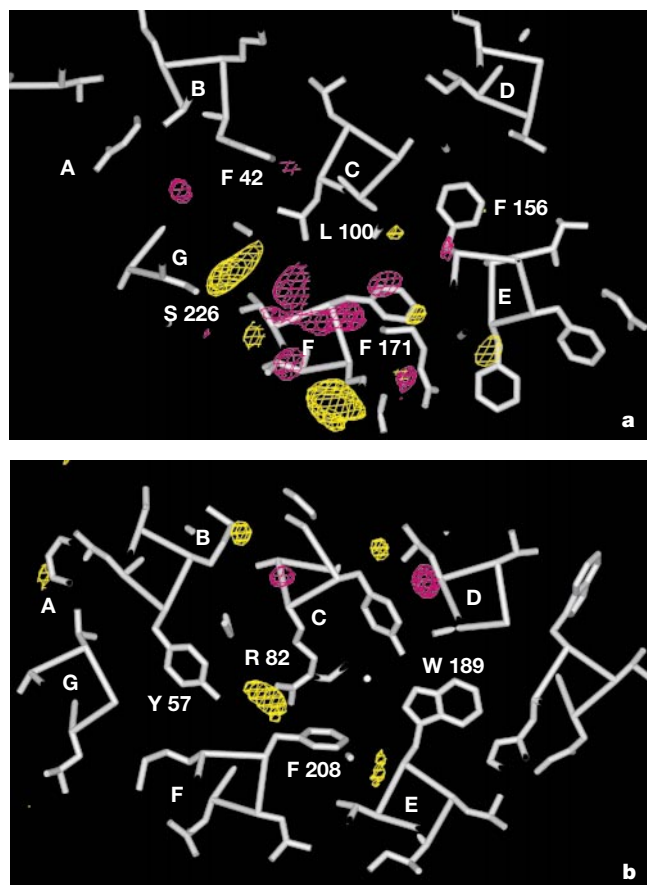


Figure 1 Three-dimensional difference density map showing structural changes in the D96G, F171C, F219L triple mutant compared with the unilluminated native wild-type bacteriorhodopsin. The view is from the cytoplasmic side along an axis perpendicular to the plane of the membrane. Diffraction patterns from two-dimensional crystals of the D96G, F171C, F219L triple mutant³ (provided by J. Tittor and D. Oesterhelt) embedded in glucose or trehalose were recorded on a Philips CM-12 electron microscope at a specimen temperature of -180°C using a CCD camera system²⁵. About 1,000 diffraction patterns were included in the initial set, from which 402 patterns were chosen by visual inspection, and merged together (merging R -factor of 17.3%) to generate a set of lattice lines covering $\sim 87\%$ of (three-dimensional) reciprocal space, with a resolution of 3.2 Å in-plane and 3.6 Å vertically. Diffraction intensities from the merged data set were scaled to those of native bacteriorhodopsin. Three-dimensional difference maps ($F_{\text{triple}} - F_{\text{native}}$) Φ_{native} were calculated using experimental amplitudes for the triple mutant (this work) and previously measured amplitudes and phases for wild-type bacteriorhodopsin²⁶. The difference densities are contoured at 3σ . Yellow, positive densities; purple, negative densities. **a**, Section close to cytoplasmic boundary where the largest structural differences are observed, with prominent features in the vicinity of helices F and G. The difference densities near helix F indicate an outward displacement of helix F in the triple mutant. **b**, Section close to extracellular boundary. The main change in this region is a positive feature localized to the vicinity of Arg 82. The differences here are considerably smaller than in the section in **a**.

reveals the global features of the conformational change. The changes are mainly localized to the cytoplasmic region, as expected from the experimental three-dimensional difference density map (Fig. 1). The outward movement of helix F is evident starting from residue Trp 182, whereas the changes in helix G are evident starting from residue Ile 222. The top (cytoplasmic) end of helix F is displaced by ~ 3.5 Å away from the centre of the protein, whereas helix G is displaced by ~ 2 Å towards a position between helix F and the centre of the protein and also becomes more ordered at its carboxy terminus. The extent of the outward tilt of helix F is about 6 degrees, with Pro 186 probably serving as a hinge residue. The large movement of helix F leads to a significant rearrangement of the EF loop, and to minor rearrangements of the uppermost residues in helix E. The movements of helices F and G also appear to cause the movement and/or disordering of a nearby lipid molecule on the cytoplasmic side (Fig. 2b).

What is the consequence of this structural change in the cytoplasmic region? Figure 3 shows the van der Waals surfaces of the cytoplasmic region around residue 96. In wild-type bacteriorhodopsin (Fig. 3a), residues Phe 42 (helix B), Leu 100 (helix C), Thr 170 (helix F), Phe 171 (helix F) and Leu 223 (helix G) effectively block the proton pathway from the cytoplasm to Asp 96. The region between Asp 96 and the Schiff base is largely non-polar, packed with bulky residues such as Leu 93 (helix C), Leu 174 (helix F) and Phe 219 (helix G). Therefore, in this 'closed', unilluminated state of the protein, the Schiff base is effectively inaccessible to protons from

the cytoplasmic region, and the pK of Asp 96 is high. In contrast, there is a clearly visible 'opening' of the upper cytoplasmic region (in the vicinity of residue 96) in the triple mutant (Fig. 3b). The bulk of the volume increase in the channel region comes from the outward movement of helix F. Despite the movement of helix G towards the central region of the channel, this movement of helix F appears to be sufficiently large to allow entry of water molecules into the channel region. In particular, the outward tilt of helix F defines a clear path for proton conduction from the cytoplasmic boundary to the Schiff base. Although minor rearrangements are observed in the orientations of side chains of residues that line the path of the proton, such as Leu 97, Leu 174 and Leu 223, inspection of the structures leaves no doubt that the global conformational change in helices F and G is the main factor contributing to the increased accessibility of the Schiff base to protons in the cytoplasm.

Recently, three other studies, also aimed at determination of the structure of the cytoplasmically open conformation of bacteriorhodopsin, have been reported. One is based on electron crystallographic analysis of the F219L mutant⁴, whereas the other two are based on analysis of X-ray diffraction patterns of three-dimensional crystals of either the D96N mutant⁵ or wild-type bacteriorhodopsin⁶. Although the analysis of the F219L mutant is at a lower vertical resolution than ours, and it used a rigid body refinement procedure that required a number of assumptions about the nature of the structural change, the results are in excellent agreement with our work. In contrast, the conclusions from the two

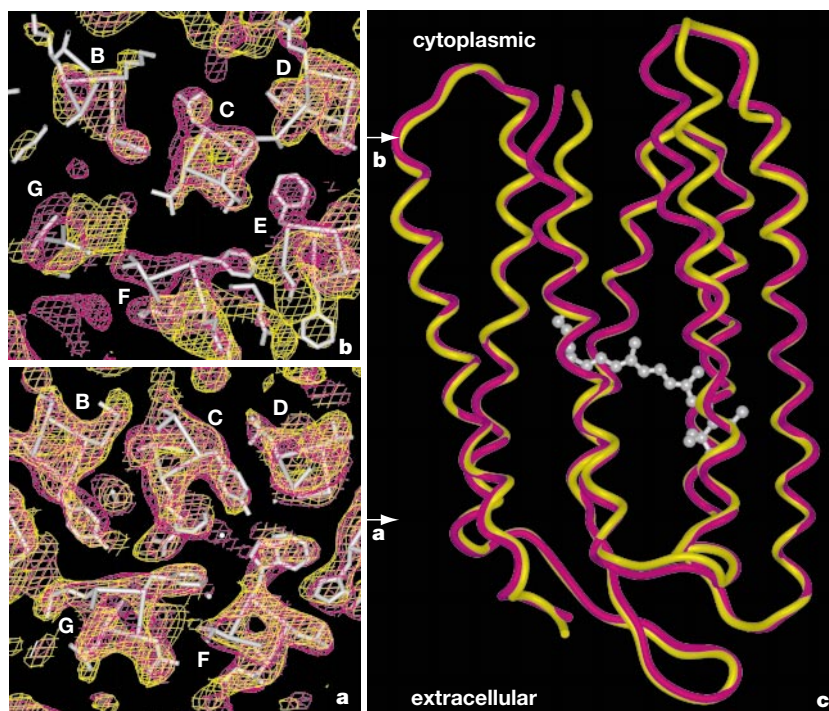


Figure 2 Comparison of atomic models for wild-type bacteriorhodopsin and the D96G, F171C, F219L triple mutant. **a, b**, Sections of σ_A weighted $2F_o - F_c$ density maps for wild-type bacteriorhodopsin (purple) and the D96G, F171C, F219L triple mutant (yellow) near the cytoplasmic (**a**) and extracellular (**b**) boundaries. The maps are superimposed on the structure of wild-type bacteriorhodopsin. Significant rearrangements of helices F and G occur in the cytoplasmic, but not in the extracellular region. **c**, Superposition of α -atoms of wild-type bacteriorhodopsin (purple) and the D96G, F171C, F219L triple mutant (yellow) illustrating differences in the cytoplasmic portions of helices F and G. Initially, a minimal starting model containing only the transmembrane regions of bacteriorhodopsin was used in a simplified least squares refinement. Coordinates from six different starting models (PDB entries 2brd, 1br, 1ap9, 1brx, 1at9 and 2at9) were tested using diffraction data sets obtained both from wild-type bacteriorhodopsin and the triple mutant. The 1br coordinates²⁸ were used as a starting model for the next stage of refinement using the

crystallography and NMR system (CNS)²⁷ suite of programs, involving simulated annealing followed by temperature factor refinement. The wild-type structure was refined to a final R -factor of 23.9% (R_{free} 31%), and the triple mutant structure was refined to a final R -factor of 27.2% (R_{free} 32.1%). The final map was tested by completely omitting from the starting model the side chains from a series of test residues such as Phe 42, Trp 86, Trp 189 and Phe 208, or various combinations of residues at the cytoplasmic ends of helix F or helix G. In each case, difference maps ($F_o - F_c$) obtained at the end of the refinement were unambiguous and clear density peaks were observed for each of the omitted regions in complete agreement with the atomic models reported here. As the diffraction data from the triple mutant and wild-type bacteriorhodopsin are completely independent of the 1br starting coordinates, these omit maps constitute stringent and objective tests of the entire refinement procedure.

X-ray crystallographic analyses disagree with each other, and with our findings in regard to the nature of the main conformational change.

Luecke *et al.*⁵ illuminated transiently thawed three-dimensional bacteriorhodopsin crystals frozen at 100 K to determine the structural changes in the photocycle of the D96N mutant (PDB entry 1C8S). They found that formation of the M intermediate is associated with increased disorder, but did not report significant differences in conformation of the polypeptide in the illuminated crystals. Because of the disorder, they were also unable to determine coordinates for residues 154–175 (the cytoplasmic sections of helices E and F, and the E–F loop), and 223–228 (the cytoplasmic end of helix G). These regions correspond almost exactly to the sites where we observe the protein conformational change.

Sass *et al.*⁶ report similar experiments with illuminated crystals of wild-type bacteriorhodopsin with an estimated occupancy of 35% of the M₂ intermediate (PDB entry 1CWQ). Their co-ordinates bear a superficial resemblance to ours in that there are rearrangements in the vicinity of helices F and G. However, their structure differs qualitatively and quantitatively from the one that we present here.

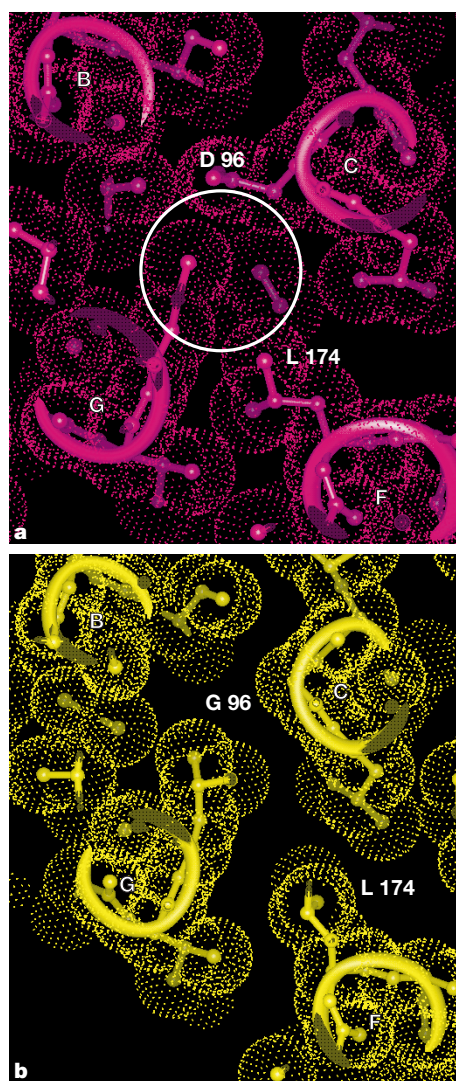


Figure 3 Representation of van der Waals surfaces at the entrance of the cytoplasmic channel in unilluminated wild-type bacteriorhodopsin and the D96G, F171C, F219L triple mutant. **a**, Wild-type; **b**, triple mutant. The four central helices lining the channel are shown, and the white circle roughly indicates the centre of the channel. The cavity between these four helices in **b** is large enough to allow the entry of one or more water molecules, increasing accessibility of the Schiff base to protons from the cytoplasm.

Importantly, where we report the largest movements within helix F (2–3 Å in the top few residues), the coordinates of Sass *et al.*⁶ contain movements of about 1 Å and are in a different direction.

One origin of the differences between these studies could be that our results are from a mutant locked in the cytoplasmically open state even before illumination, whereas the others involve the use of illuminated crystals. This possibility can be excluded because comparison of projection difference maps at 3.5 Å resolution reveals that the structure of the triple mutant must be the same as that observed upon illumination of either wild-type bacteriorhodopsin or the D96N mutant under physiological conditions³. To stringently test this similarity, we calculated a projection difference Fourier map using the atomic coordinates of structures for the D96G, F171C, F219L triple mutant and wild-type bacteriorhodopsin (Fig. 4). This map has all of the features previously observed in the experimental projection difference maps obtained by illumination of either wild-type bacteriorhodopsin or the D96N mutant under fully hydrated conditions in the native bilayer environment, using neutron, X-ray or electron diffraction methods^{3,7–10}. Similar test maps calculated using the 1C8S (ref. 5) or 1CWQ (ref. 6) coordinates do not reproduce the principal features of the previously published experimental difference maps (data not shown).

The contradictory findings from the electron and X-ray crystallographic experiments cannot be attributed to intrinsic differences between electron and X-ray crystallographic methods, as a 2 Å

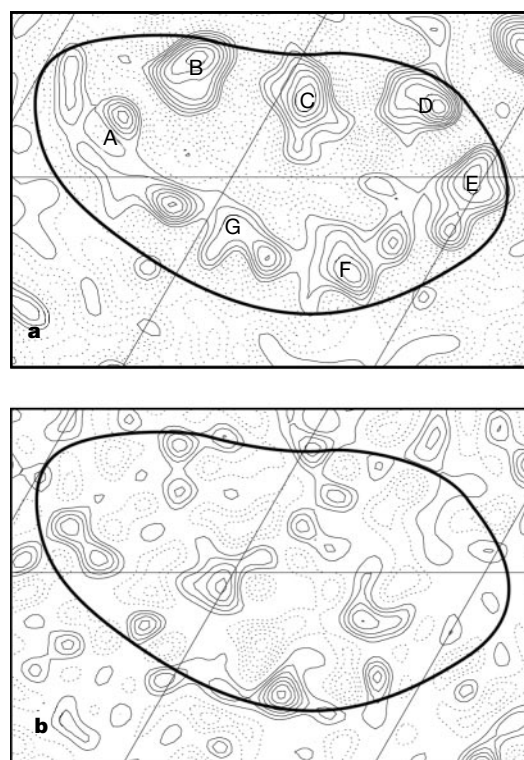


Figure 4 Analysis of the nature of the protein conformational change using projection Fourier maps. **a**, Two-dimensional projection Fourier map of wild-type bacteriorhodopsin at 3.5 Å resolution, identifying locations of the seven transmembrane helices. The grid bars are spaced at 20.8 Å, which is one-third the length of the unit cell. The solid line shows the approximate boundary of individual molecules in the lattice. **b**, Projection difference Fourier map calculated using amplitudes from the coordinates for the D96G, F171C, F219L triple mutant and amplitudes and phases from coordinates for wild-type bacteriorhodopsin. The contour interval is one-eighth of that used in **a**. This map reproduces features observed in experimentally determined projection difference maps that describe structural changes in illuminated crystals of wild-type bacteriorhodopsin or the D96N mutant³.

movement of the cytoplasmic end of helix G was directly observed in the M state of a mercury-labelled cysteine mutant (D96N/I222C) by X-ray powder diffraction¹¹ in agreement with our results. We conclude that despite the higher resolution (2.0 Å) of the structures derived by X-ray crystallography, the main structural change is

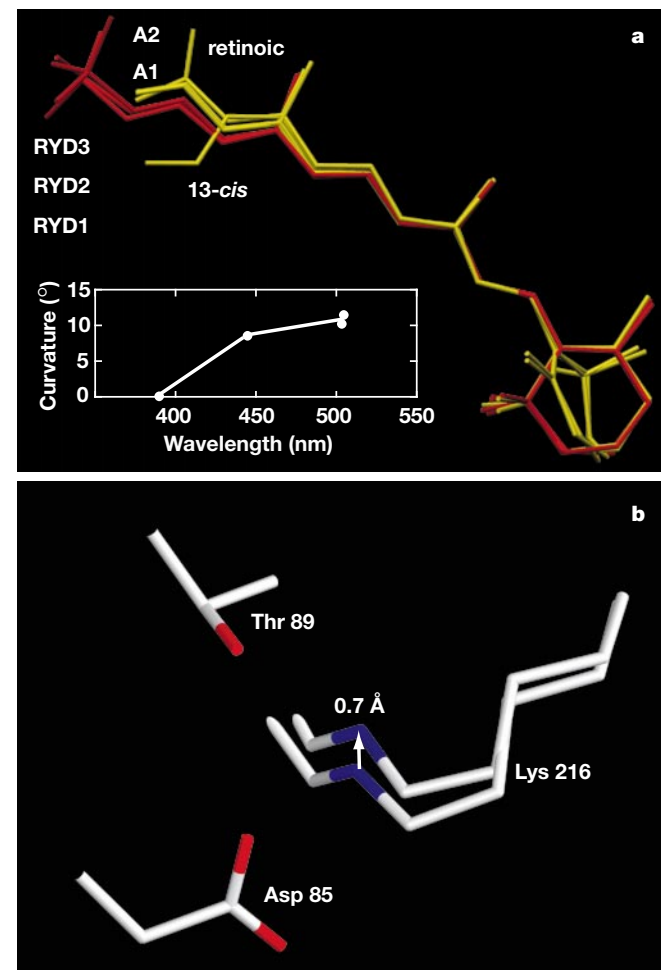


Figure 5 Observed conformations of retinal derivatives and proposed displacement of Schiff base on deprotonation. **a**, Conformation of seven retinal, retinoic acid or retinylidene derivatives determined by X-ray crystal structure analysis^{13–18}. The molecules have been superimposed by alignment of atoms C7 to C9 and C19. The seven structures are retinaldehyde in the all-*trans* (A1) and 13-*cis* (13-*cis*) conformation; 3,4-didehydro retinaldehyde (A2); all-*trans* retinoic acid (retinoic); *N*-methyl-*N*-phenyl retinylideneiminium (RYD1) and two different *t*-butyl-retinylideneiminium salts (RYD2, RYD3). Except for 13-*cis* A1, the rest have been crystallized in the all-*trans* configuration. The first four (yellow) are uncharged and the rest (red) are positively charged. Inset, curvature of the polyene backbone of retinal derivatives plotted against the wavelength of the absorption maximum, λ_{max} . There is a trend towards greater curvature with protonation and a higher λ_{max} . The extra curvature in the positively charged retinylidenes is accompanied by greater delocalization of the charge as judged by the reduction in the average difference in bond length between single and double bonds along the length of the polyene chain (data not shown). The extra curvature in RYD1, the most curved of the structures, averages about 1.0 degree per bond angle and seems to be evenly distributed along the length of the molecule. **b**, A model for displacement of the nitrogen in the retinylidene Schiff base upon deprotonation during the photocycle. Average structures for protonated and unprotonated states of retinal were based on the structures of the compounds shown in **a**. Bond lengths and angles derived from these were used in a least squares refinement to obtain models for the geometry of the Schiff base region in the L and M intermediates. The reduced curvature of retinal in the M intermediate displaces the Schiff base nitrogen towards the cytoplasmic side by about 0.7 Å, connecting it to the cytoplasmic proton channel and reducing accessibility to the extracellular proton channel.

either inhibited or muted in the three-dimensional crystals. However, the possibility cannot be excluded that selected features of the coordinates derived by X-ray crystallographic refinement bear a resemblance to one or more of the early intermediates in the photocycle (for example, L, M₁) that precede the main conformational change.

Does the structural change in the cytoplasmic region completely account for the change in accessibility of the Schiff base after release of the proton to the extracellular medium? There are no further significant changes in global protein conformation in the D96G, F171C, F219L triple mutant upon illumination³. However, biochemical analyses have shown that this mutant can also pump protons against a transmembrane potential gradient of about one half of that of wild-type bacteriorhodopsin (J. Tittor and D. Oesterhelt, personal communication). Light-driven isomerization alone cannot provide the necessary switching action, as the retinylidene Schiff base retains its 13-*cis*, 15-*anti* configuration at both the time of proton release and uptake¹². Therefore, under conditions when the large protein structural change is present throughout the photocycle, there must be further subtle rearrangements of retinal and/or the protein that are responsible for the change in accessibility of the Schiff base.

As the retinylidene moiety is intimately involved in the proton release and uptake events, we have examined high-resolution X-ray crystal structures of seven retinylidene compounds^{13–18} to look for clues that might explain its role in ensuring the vectoriality of proton pumping. Inspection of these structures (Fig. 5a) reveals two important features: first, the intrinsic curvature of the polyene chain is greater in the positively charged molecules than in the neutral molecules; and second, the overall extent of the increased curvature of the positively charged molecules correlates with the wavelength of maximal absorption (Table 1; Fig. 5a, inset). This extra ‘backbone’ curvature is an intrinsic feature of all charged retinal molecules, leading to the conclusion that a similar difference in polyene curvature must also occur between deprotonated and protonated Schiff base intermediates in the bacteriorhodopsin photocycle.

The difference in curvature of the polyene chain between protonated and unprotonated states of the Schiff base has a profound consequence on the local geometry of the Schiff base region during the photocycle. The β -ionone ring end of retinal is held in place by steric interactions with the protein, so the change in curvature will have the largest effect at the Schiff base nitrogen, displacing it towards the cytoplasmic side. The extent of this displacement must be at least 0.7 Å, which is the curvature difference between *N*-methyl-*N*-phenyl retinylideneiminium (RYD1) and the neutral retinal compounds at the position corresponding to the Schiff base nitrogen (Fig. 5b). This value is probably underestimated, because the λ_{max} of RYD1 is 505 nm, as compared to 550 nm for the protonated Schiff base intermediates (L and N) occurring before and after the deprotonated M intermediate in the photocycle. The predicted movement of the Schiff base nitrogen is consistent with spectroscopic measurements that find an increase of about 7° in inclination of the chromophore with respect to the membrane normal¹⁹, and an increased distance between Asp 212 and the C14 atom of retinal upon illumination²⁰. Importantly, this ‘flexional’ movement of retinal has the correct polarity to break the connection of the Schiff base with the extracellular proton release pathway, and to orientate it towards the cytoplasmic proton uptake pathway, providing a mechanism to ensure the vectoriality of proton pumping.

The combination of the change in retinal curvature upon Schiff base deprotonation and the protein structural changes in the cytoplasmic region provide the necessary elements to provide a detailed explanation for the key steps in the photocycle which underlie vectorial proton pumping by bacteriorhodopsin. Upon light absorption, the configuration of retinal is changed to the 13-*cis*, 15-*anti* geometry, and the Schiff base proton is probably

Table 1 Spectroscopic and geometric characterization of retinals with known structures

Structure	Charge state	λ_{max} (nm)*	Angle between bonds C9–C19 and C13–C20	Extra curvature C9 to C13	Net increase in overall curvature†
retinalA1	0	~390	11.9	0.3	~0
retinalA2	0	~390	11.6	0.0	~0
retinoic acid	0	~390	12.1	0.5	~0
13- <i>cis</i> retinal	0	~390	10.7	-0.9	~0
Retinylidene iminium 1	+1	505	18.0	6.5	(12.9, 10.0) 11.4
Retinylidene iminium 2	+1	504	17.2	5.6	(11.2, 9.0) 10.1
Retinylidene iminium 3	+1	445	16.0	4.4	(8.8, 8.1) 8.5

* λ_{max} is not available for uncharged retinal derivatives in their crystalline form, so 390 nm is an estimate. For the retinylideneiminium compounds 1, 2 and 3, λ_{max} are for the solid (crystalline) state¹³.

† Calculated in two ways: the first number represents simple doubling of the extra C9 to C13 curvature given in the previous column; the second number is the overall curvature between C6–C7 and C14–C15 in the retinylideneiminium structures, but for the uncharged retinal derivatives, corrections were applied to eliminate end effects due to the 6-*s-cis* information in A1 and A2, the 13-*cis* bond in 13-*cis* retinal and the carboxyl moiety in retinoic acid. The third number is the average of the two calculations and is plotted in the inset to Fig. 5a.

transferred to Asp 85 by successive transfer through hydrogen bonds between the Schiff base and Thr 89, and between Thr 89 and Asp 85. The loss in electrostatic interaction between the Schiff base and Asp 85 is known to be important in driving the transition to the cytoplasmically open state²¹. The upward movement of retinal further aids this transition, as the C13 and C9 methyl groups are in direct van der Waals contact with Trp 182 in helix F, a residue located one turn away from Pro 186 that serves as the hinge residue for the tilt of helix F. In the D96G, F171C, F219L triple mutant, the destabilizing effects of the three mutations are strong enough to override modulations from changes in retinal isomerization, retinal curvature and electrostatic interactions near the Schiff base that are normally required for shifting the equilibrium during the photocycle of wild-type bacteriorhodopsin.

Two other types of seven-helix retinal proteins closely related to bacteriorhodopsin are found in the cell membranes of *H. salinarum*. Halorhodopsin, with a relative molecular mass of 28,000 (M_r 28K), pumps chloride ions into the cytoplasm from the extracellular medium upon illumination²². Two sensory rhodopsins (M_r ~26K) direct phototaxis of the bacterium towards green and away from blue light, respectively²³. In these sensory rhodopsins, light absorption activates an associated dimeric transducer protein that has two pairs of transmembrane helices and is homologous to a large family of transducer proteins that mediate chemotaxis in bacteria. In turn, these transducer proteins activate a histidine kinase, setting in motion events that change the direction of rotation of the flagellar motor.

Given the high conservation of residues in the retinal binding pocket and inside the protein, it is probable that there is a light-driven opening of the cytoplasmic channel in the halobacterial rhodopsins similar to that for bacteriorhodopsin. In halorhodopsin, this opening would provide an exit pathway for the chloride ion from the Schiff base region to the cytoplasm after the initial light-driven retinal isomerization. In the sensory rhodopsins²³, helices F and G are likely to interact with two helices of the transducer to trigger activation similarly to that induced by ligand binding to chemotaxis receptor dimers. In visual rhodopsin, the G-protein-coupled receptor, the light-driven structural change of 11-*cis* into all-*trans* retinal is also transmitted to the extramembranous regions through the seven α -helices that surround retinal, activating the cytoplasmic G-protein, transducin. Interestingly, spectroscopic and mutagenesis studies indicate that structural changes in the cytoplasmic ends of helices E and F, and the E–F loop region are especially important in light-induced G-protein activation²⁴. Although it is certain that there will be differences in detail, the type of helix movements described here for bacteriorhodopsin might also lie at the heart of the mechanism of receptor activation in visual rhodopsin and related G-protein-coupled receptors. □

- Subramaniam, S. The structure of bacteriorhodopsin: an emerging consensus. *Curr. Opin. Struct. Biol.* **9**, 462–468 (1999).
- Subramaniam, S. *et al.* Protein conformational changes in the bacteriorhodopsin photocycle. *J. Mol. Biol.* **287**, 145–161 (1999).
- Vonck, J. Structure of the bacteriorhodopsin mutant F219L N intermediate revealed by electron crystallography. *EMBO J.* **19**, 2152–2160 (2000).
- Luecke, H., Schobert, B., Richter, H. -T., Cartailler, J. -P., & Lanyi, J. K. Structural changes in bacteriorhodopsin during ion transport at 2 Å resolution. *Science* **286**, 255–260 (1999).
- Sass, H. -J. *et al.* Structural alterations for proton translocation in the M state of wild-type bacteriorhodopsin. *Nature* **406**, 649–653 (2000).
- Dencher, N. A., Dresselhaus, D., Zaccari, G. & Büldt, G. Structural changes in bacteriorhodopsin during proton translocation revealed by neutron diffraction. *Proc. Natl Acad. Sci. USA* **86**, 7876–7879 (1989).
- Koch, M. H. J. *et al.* Time-resolved X-ray diffraction study of structural changes associated with the photocycle of bacteriorhodopsin. *EMBO J.* **10**, 521–526 (1991).
- Kamikubo, H. *et al.* The last phase of the reprotonation switch in bacteriorhodopsin: The transition between the M-type and the N-type protein conformation depends on hydration. *Biochemistry* **36**, 12282–12287 (1996).
- Subramaniam, S., Gerstein, M., Oesterheld, D. & Henderson, R. Electron diffraction analysis of structural changes in the photocycle of bacteriorhodopsin. *EMBO J.* **12**, 1–8 (1993).
- Oka, T. *et al.* Conformational change of helix G in the bacteriorhodopsin photocycle: Investigation with heavy atom labeling and X-ray diffraction. *Biophys. J.* **76**, 1018–1023 (1999).
- Fodor, S. P. *et al.* Chromophore structure in bacteriorhodopsin's N intermediate: implications for the proton pumping mechanism. *Biochemistry* **27**, 7097–7101 (1988).
- Elia, G. R., Childs, R. F., Britten, J. F., Yang, D. S. C. & Santarsiero, B. D. Structure and wavelength modification in retinylidene iminium salts. *Can. J. Chem.* **74**, 591–601 (1996).
- Santarsiero, B. D., James, M. N. G., Mahendran, M. & Childs, R. F. The crystal structure of N-methyl-N-phenyl-retinylideneiminium perchlorate: a structural model for the bacteriorhodopsin chromophore. *J. Am. Chem. Soc.* **112**, 9416–9418 (1990).
- Hamanaka, T., Mitsui, T., Ashida, T. & Kakudo, M. The crystal structure of all-*trans* retinal. *Acta Crystallogr. B* **28**, 214–222 (1972).
- Stam, C. H. The crystal structure of a monoclinic modification and the refinement of a triclinic modification of vitamin A acid (retinoic acid), C₂₀H₂₈O₂. *Acta Crystallogr. B* **28**, 2936–2945 (1972).
- Simmons, C. J., Liu, R. S. H., Denny, M. & Seff, K. The crystal structure of 13-*cis*-retinal. The molecular structures of its 6-*s-cis* and 6-*s-trans* conformers. *Acta Crystallogr. B* **37**, 2197–2205 (1981).
- Simmons, C. J., Asato, A. E. & Liu, R. S. H. Structure of All-*trans*-3,4-didehydroretinal (retinal). *Acta Crystallogr. C* **42**, 711–715 (1986).
- Moltke, S. *et al.* The angles between the C-1, C-5, and C-9-methyl bonds of the retinylidene chromophore and the membrane normal increase in the M-intermediate of bacteriorhodopsin: Direct determination with solid-state H-2 NMR. *Biochemistry* **38**, 11762–11772 (1999).
- Griffiths, J. M. *et al.* Structural investigation of the active site in bacteriorhodopsin: Geometric constraints on the roles of Asp-85 and Asp-212 in the proton pumping mechanism from solid-state NMR. *Biochemistry* **39**, 362–371 (2000).
- Brown, L. S. *et al.* A local electrostatic change is the cause of the large scale protein conformation shift in bacteriorhodopsin. *Proc. Natl Acad. Sci. USA* **94**, 5040–5044 (1997).
- Schobert, B. & Lanyi, J. K. Halorhodopsin is a light-driven chloride pump. *J. Biol. Chem.* **257**, 306–313 (1982).
- Hoff, W. D., Jung, K. H. & Spudich, J. L. Molecular mechanism of photosignaling by archaeal sensory rhodopsins. *Ann. Rev. Biophys. Biomol. Struct.* **26**, 223–258 (1997).
- Altenbach, C. *et al.* Structural features and light-dependent changes in the cytoplasmic interhelical E-F loop region of rhodopsin: a site-directed spin-labeling study. *Biochemistry* **35**, 12470–12478 (1996).
- Faruqi, A. R., Henderson, R. & Subramaniam, S. Cooled CCD detector with tapered fibre optics for recording electron diffraction patterns. *Ultramicroscopy* **75**, 235–250 (1999).
- Grigorieff, N., Ceska, T. A., Downing, K. H., Baldwin, J. M. & Henderson, R. Electron-crystallographic refinement of the structure of bacteriorhodopsin. *J. Mol. Biol.* **259**, 393–421 (1996).
- Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Essen, L. O., Siebert, R., Lehmann, W. D. & Oesterheld, D. Lipid patches in membrane protein oligomers: Crystal structure of the bacteriorhodopsin-lipid complex. *Proc. Natl Acad. Sci. USA* **95**, 11673–11677 (1998).

Correspondence and requests for materials should be addressed to S.S. (e-mail: ss1@nih.gov). The coordinates for the structures reported here are deposited in the Protein Data Base (accession numbers 1FBB for wild-type bacteriorhodopsin and 1FBK for the mutant).

Received 14 April; accepted 12 June 2000.

- Henderson, R. *et al.* Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* **213**, 899–929 (1990).

Microarrays on the slide

New gadgets, including some of the latest in microarray technology.

Coated microscope slides

Xenopore

www.xenopore.com

Silanated slides, and others

Glass slides created by reacting the surface with 3-aminopropyl silane have a uniform and consistent amino surface, according to Xenopore, who offer the slides with either plain or frosted ends. Silylated slides with a reactive aldehyde surface for binding amino-labelled DNA and RNA probes, proteins and peptides are also available, as is an epoxy-coated slide that binds to primary amino groups. In addition, the company supplies streptavidin-coated slides for binding biotin-labelled oligos, nickel chelate slides for His-tagged molecules, maleimide slides for sulphhydryls, Con A for carbohydrates and polylysine for cell binding.

Reader Service No. 100

StepVial system

J. G. Finneran Associates www.jgfinneran.com

Two-piece sample vial for autosamplers

This sample vial has a limited-volume insert with a choice of closure designs. The insert can be placed precisely in the centre of the vial neck, avoiding accidental jamming of the autosampler needle, and does not require a spring assembly to prevent 'bottoming out'. An enlarged diameter neck opening on both the vial and the insert provides a larger target area for the autosampler and facilitates manual pipetting.

Reader Service No. 101

Custom Atlas array service

Clontech

www.clontech.com

Gene expression profiling

This service provides customized arrays containing a selected set of genes either provided by the customer or taken from its Atlas gene collection. Arrays of up to 1,176 cDNAs each can be produced, and fragments can be arrayed on either nylon or glass. The company will also carry out experiments in-house using either pre-made or custom arrays, with image files analysed using AtlasImage software.

Reader Service No. 102

SPBIO Microarray Station

Hitachi Genetic Systems

www.mirabio.com

SPBIO marks the spot

The SPBIO Microarray Spotting Station from Hitachi Genetic/MiraiBio Inc. is a fully enclosed system which uses a proprietary pin design built for spot reproducibility and precision. The pin tip can spot on glass slides or membranes, and spot pitch can be adjusted



Pin number: the SPBIO Spotting Station.

from 0.2 to 1.00 mm in 0.01-mm increments. The SPBIO accepts a variety of DNA samples, including oligo and cDNA. Different pin sizes are available. The SPBIO's compact design accommodates 96- or 384-well plates with a cover to keep samples clean. Up to 48 chips can be prepared under the same setting.

Reader Service No. 103

GeneSpring 3.0

Silicon Genetics

www.sigenetics.com

Branch out with new software

This software package for genomic expression analysis is optimized for both MicroArray and Affymetrix experimental data sets. New features include an improved data import wizard, an enhanced overlay feature for KEGG custom pathways, multidimensional tree-building capabilities, linked tiled windows and bookmark support. The program is compatible with both PCs and Macs.

Reader Service No. 104

Open Access Centrifugal Evaporator

Genevac

www.genevac.com

Fractionally better

Designed to form an integral part of an LC-MS purification system for medicinal chemistry, this evaporator allows collected fractions to be inserted and retrieved without batching, speeding up the processing of purified fractions. New fractions can be inserted into empty spaces in the racks within the evaporator or replace dry samples that are



In a spin: centrifugal evaporation from Genevac.

ready for removal. Fractions can be identified easily at all stages of the process.

Reader Service No. 105

PixSys 5500XL

Cartesian Technologies

www.cartesiantech.com

Integrated printer and plate loader

Capable of processing up to 50 slides, the PixSys 5500XL is an integrated system consisting of a printer, plate loader and humidity chamber for automated printing of microarrays. Plates are transferred to and from the printing workstation using a pick-and-place robotic arm. Pins can be easily arranged for multiple configurations and source plate types. There is a vacuum slide nest for each slide. The system can handle spot sizes from 120 to 200 µm, or up to about 23,000 spots per slide, depending on application.

Reader Service No. 106

AlphaEase FluoChem 2.0

Alpha Innotech

www.alphainnotech.com

Download it for free

This software release for low-light imaging in macroarray, fluorescence, chemiluminescence and visible imaging offers new features, including a 'magic wand' that thresholds bands and spots and provides corresponding densitometry information. Background subtraction options in the 1D multi-lane analysis function include 'rubber band', 'minimum profile', 'rolling disk' and 'valley-to-valley'. The MovieMode image capture feature utilizes image stacking as it acquires images. AutoLane and AutoSpot allow detection of lanes and bands in a gel image and 2D detection and quantification of spots. Contrast can be adjusted for low-light chemilumi-

new on the market

nescence imaging using 'ChemiDisplay'. Alpha Innotech customers can download AlphaEase FC 2.0 free from the company's website.

Reader Service No. 107

Micromax System I-direct

NEN www.nen.com

Simple and direct microarray technology

The Micromax system uses direct fluorescence cDNA labelling with proprietary cyanine 3- and cyanine 5-labelled nucleotides. Since components and processing steps are minimized, the system is particularly suitable for laboratories just beginning to establish and validate microarray applications. The company also offers a variety of products for use in high-throughput, differential gene expression research.

Reader Service No. 108

FAST Slides

Scleicher & Schuell www.s-und-s.de

Microhybridization on a membrane

This new kit has been developed with a view to producing high-quality consistent fluorescent data from membrane-based microarrays. The FAST Slides MicroHybridization kit comprises slides with a microporous polymeric surface, to give a consistent surface, buffers optimized for fluorescence detection and disposable hybridization chambers. FAST slides are produced in collaboration with race Bio Labs of Bend, Oregon.

Reader Service No. 109

Model 400 Hybridization Incubator

Robbins www.robsci.com

Rock and rotate

This unit, designed for processing membrane macroarrays, has a tri-directional rotator and a high-energy heating element and fan that can quickly raise the temperature of the chamber to any desired value (maximum 90 °C). Temperature can be regulated to $\pm 0.1^\circ\text{C}$. For greater numbers of blots, a max-rotator is available that can be probed with up to 80 dot membranes per session. A rocker platform for gentle movement of solutions in dishes, trays and tissue culture flasks is also available.

Reader Service No. 110

Assist

ThermoQuest www.thermoquest.com

Automated benchtop microplate handling

This device automatically moves individual microplates from an input rack onto the plate carrier and triggers the reader application, after which the plate is retrieved from the reader and placed on an output rack. The basic system can sequentially load up to twenty 96-well plates, with an upgrade available to increase this to eight. Plates with lids



Help: the Assist system boasts high capacity.

can also be used. Data for each plate can be produced using the company's software.

Reader Service No. 111

ImaGene

Genetic Research www.gri.co.uk

Software to capture microarray data

This program from Genetic Research Information is powerful enough to process a three

million pixel image with over 10,000 spots in under two minutes, but, say the manufacturer, is still simple to use. It is suitable for processing radioactive as well as fluorescent labels, an unlimited number of spots, grids of any size and irregularly arrayed data points. Applications include capture of gene expression microarray and high-density membrane data.

Reader Service No. 112

MicroTek AW, AWS and AWM

Bio-tek Kontron www.biotek-kontron.com

A trio of microplate washers

These washers are designed for high-throughput screening. The MicroTek AW for standard 96-well microplates features computer control via RS232 port, on-board programmable software for magnetic and standard assays, and aspiration and dispense height adjustment. The AWS model for 96- and 384-well washing allows independent adjustment of dispense and aspirate portions of the manifold and is robot-compatible. For processing magnetic bead assays, the AWM model will wash standard 96-well and magnetic particle assays.

Reader Service No. 113

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENTS

TREVIGEN

new

PARP

**Poly (ADP-ribose)
Polymerase Reagents**

- PARP inhibition assay
- Biotinylated NAD
- Recombinant human PARP
- Anti PARP and PAR

www.trevigen.com

US and Canada: 1 800 873 8443
TEL: 301 216 2800 • FAX: 301 216 2801

READER ENQUIRY NO. 43

mtDNA EXTRACTOR KITS

NEW

- Easy & efficient isolation of mtDNA from whole blood, cell culture or tissue
- Results in ~90 Minutes
- No Hazardous Reagents
- Works even in the presence of anticoagulants

Wako

Wako USA Wako GmbH

877-714-1921 (49)02131 3110

www.wako-chem.co.jp

READER ENQUIRY NO. 38

Look no further

for the best value in

Oligo Synthesis

Competitive pricing
Routine 24 hour dispatch • Every Oligo QC verified by PAGE
Rigorous batch QC by PCR and Mass Spec
Web ordering and tracking • Local Freephone numbers
A great deal in any language!!

SIGMA
GENOSYS

(+44) (0) 1223 839000 • fax (+44) (0) 1223 839200
email: info@sigma-genosys.co.uk • www.sigma-genosys.co.uk

READER ENQUIRY NO. 33

BACs • YACs • PACs cDNAs

Research Genetics provides clones and colony membranes for a number of human, mouse and rat libraries as well as a custom screening service for the libraries.

Research Genetics
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016

Check out our homepage at
<http://www.resgen.com>

READER ENQUIRY NO. 40